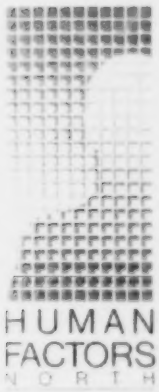
DRDC Toronto CR 2003-047

VISUAL COMPLEXITY: A REVIEW

by
Don C. Donderi, PhD
Human Factors North Inc.
118 Baldwin Street
Toronto, Ontario
M5T 1L6



Project Manager - D.C. Donderi, PhD
(514) 398-6130
Submitted in partial fulfillment of:
PWGSC Contract No. W7711-017758/001/TOR

On behalf of
DEPARTMENT OF NATIONAL DEFENCE
As represented by:
Defence Research and Development Canada (DRDC)
1133 Sheppard Avenue West
Toronto, Ontario, Canada
M3M 3B9

DRDC Scientific Authority
Sharon McFadden
Phone: (416) 635-2189
March 4, 2003

The scientific or technical validity of this Contract Report is entirely the responsibility of the contractor and the contents do not necessarily have the approval or endorsement of Defence R&D Canada.

104-06919/

© HER MAJESTY THE QUEEN AS REPRESENTED BY THE MINISTER OF NATIONAL DEFENCE, 2003
© SA MAJESTÉ LA REINE, REPRÉSENTÉE PAR LE MINISTRE DE LA DÉFENSE NATIONALE, 2003

As represented by the Minister of National Defence

VISUAL COMPLEXITY: A REVIEW

LA COMPLEXITÉ VISUELLE: UNE RÉCAPITULATION

Abstract

The definition and measurement of the complexity of visual forms and visual arrays is reviewed, starting with the contributions of gestalt psychology at the beginning of the twentieth century and ending with current research in algorithmic information theory, perceptual learning theory, and neural circuit theory. Complexity and probability are shown to be linked through their reciprocal relationship in algorithmic information theory, which is shown in turn to be consistent with important aspects of perceptual learning theory and neural circuit theory. The emphasis is on applications to visual displays.

Résumé

On passe en revue la définition et la mesure de la complexité des formes visuelles et des groupements visuels, depuis les contributions du gestaltisme au début du 20^{e} siècle jusqu'aux recherches actuelles sur la théorie de l'information algorithmique, la théorie de l'apprentissage perceptuel et la théorie du circuit neuronal. Il est démontré que la complexité et la probabilité sont interreliées dans le cadre de la théorie de l'information algorithmique, qui s'avère quant à elle compatible avec d'importants aspects de la théorie de l'apprentissage perceptuel et de la théorie du circuit neuronal. L'accent est mis sur les applications dans le contexte de la visualisation.

Executive Summary

The idea of visual complexity is followed from its origin in the gestalt psychology of the early 1900s to its present place in neural circuit theory, algorithmic information theory, and perceptual learning theory. The use of this idea in understanding the perception of single visual forms, as well as arrays of visual forms, is considered, and a modern interpretation of display complexity based on algorithmic information theory is discussed.

Two different approaches to understanding visual perception developed in the late nineteenth and early twentieth centuries. One approach emphasized the analysis of the visual stimulus into sensory units whose perceptual effects could be separately measured. With this approach, the laws of perception were associationistic combination rules for the sensory units. The other approach, gestalt psychology, was based on evidence that the perceptual properties of many stimuli could not be derived by analyzing their separate parts. Instead, the laws of visual perception were rules that specified how the entire visual sensory input could be organized into a percept. Decades of research in this tradition refined the basic gestalt idea into what is called the 'minimum principle', which was that a visual percept will be as simple as the sensory input allows it to be.

Defining and measuring 'Visual simplicity' and its opposite 'visual complexity' became a main goal of psychologists working in the gestalt tradition. Claude Shannon's theory of communication became a main tool for perception researchers, who explored the idea that visual forms could be considered as particular instances of a message drawn from a pool of messages, and that rules governing the independence (or lack of it) among the messages in the pool would affect the perception of, and behavior involving, the resulting visual forms. The concepts of form dimensions, dimensional redundancy, and the separability and integrality of perceptual dimensions grew out of this approach.

Another approach taken was to try to find a 'code' of symbols that could represent a visual form, and to equate the complexity of the visual form with the length of the code required to represent the form. It turned out that there was a close relationship between the code length required to specify a visual form and the communication theory conditions established to generate that form. This approach ultimately led to the application of algorithmic information theory (AIT) to visual perception. In AIT, the complexity of any bit string (including the bit string defining a visual form) can be measured by the length of the bit string required by a Turing machine to generate the bit string corresponding to the visual form, and the probability of that form is an inverse function of the length of the bit string required to generate the image string.

A practical analogue of the AIT definition of array complexity was found in data compression algorithms, and the possible use of a data compression algorithm to generate a measure of visual complexity is discussed.

Studies of visual search through visual textures and arrays of visual forms demonstrated that some search tasks were independent of the number of items or apparent complexity of the array, while other search tasks took time in proportion to the complexity of the array. It appears that some of the sensory inputs to the visual system are relatively easy to distinguish from one another (so-called visual pop-out), while other inputs, in appearance not much more complicated than the easier inputs, were hard to discriminate. The concept of linear separability of search background and target forms explains some of the difference: if targets and background are separated by a linear partition in the display sensory space, the task is easy; if not, the task is hard.

A convergence of ideas in AIT, perceptual learning theory and neural circuit theory has led to the following synthesis: according to perceptual learning theory, a simple visual form or array generates activity in higher-order levels of the visual system (beyond area 17, all the way to medial temporal cortex), while a complex form or array generates activity from these more central areas all the way out to the more peripheral sensory input areas of the visual system. In general: simple forms activate central visual areas; more complex forms activate central and peripheral visual areas. Simple forms have shorter AIT computational codes, which corresponds to their activation of fewer and less specific cortical sites; complex forms have longer AIT computational codes, corresponding to their activation of more and more specific cortical sites. Neural circuit theories converge on an explanation of visual perception that involves multiple layers of self-reinforcing neural circuits, each layer corresponding to a higher level of generality. Simple forms

selectively activate higher-order circuits without specifying which lower-order circuits are active; more complex forms activate both specific higher-level and lower- level forms. The application of these and other principles of display organization to understanding performance based on visual displays is discussed.

Sommaire

On passe en revue la notion de complexité visuelle depuis son origine dans la théorie psychologique du gestaltisme au début du 20^{e} siècle jusqu'à son actuelle expression dans la théorie du circuit neuronal, la théorie de l'information algorithmique et la théorie de l'apprentissage perceptuel. On examine le recours à ce concept pour parvenir à comprendre la perception de formes visuelles isolées ainsi que de groupements de formes visuelles, et on analyse une interprétation moderne de la complexité de la visualisation fondée sur la théorie de l'information algorithmique.

Deux modes d'étude de la perception visuelle sont apparus à la fin du 19^{e} et au début du 20^{e} siècle. Le premier consistait essentiellement à analyser le stimulus visuel en unités sensorielles dont les effets perceptuels pouvaient être mesurés séparément. Dans ce contexte, les lois de la perception étaient des règles de combinaison associationniste des unités sensorielles. Le second mode, soit celui du gestaltisme, reposait sur des preuves selon lesquelles les propriétés perceptuelles de nombreux stimuli ne pouvaient être déterminées par l'analyse de leurs composantes. Les lois de la perception visuelle étaient donc des règles précisant le mode d'organisation de l'ensemble de l'input visuel en une image mentale. Des décennies de recherche dans cette voie ont permis de perfectionner la notion gestaltienne fondamentale et d'en tirer le « principe minimal », selon lequel l'image mentale visuelle conserve le maximum de simplicité que permet l'input sensoriel.

La définition et la mesure de la « simplicité visuelle » et de son contraire, la « complexité visuelle », sont devenues le but premier des psychologues de l'école gestaltienne. La théorie de la communication de Claude Shannon est devenue l'outil principal des chercheurs spécialistes de la perception, qui ont exploré la possibilité que les formes visuelles soient considérées comme des manifestations particulières d'un message tiré d'une réserve ce messages et que les règles régissant l'indépendance (ou la dépendance) des messages de la réserve les uns par rapport aux autres aient une incidence sur la perception des formes visuelles ainsi produites et sur le comportement associé à ces formes. Les concepts des dimensions des formes, de la redondance des dimensions ainsi que de la séparabilité et de l'intégralité des dimensions perceptuelles sont issus de cette approche.

Une autre approche a consisté à tenter de trouver un « code » de symboles pouvant représenter une forme visuelle et d'associer la complexité de la forme visuelle à la longueur du code représentant la forme. Il s'est avéré qu'il existait un lien étroit entre la longueur du code représentant une forme visuelle et les condition de la théorie de la communication établies pour produire cette forme. Cette approche a finalement abouti à l'application de la théorie de l'information algorithmique (TIA) à la perception visuelle. Selon cette théorie, la complexité de toute chaîne de bits (dont la chaîne de bits

définissant une forme visuelle) peut être mesurée en fonction de la longueur de la chaîne de bits que requiert la machine de Turing pour produire la chaîne correspondant à la forme visuelle, et la probabilité de cette forme est une fonction inverse de la longueur de la chaîne de bits requise pour produire la chaîne de l'image.

Les algorithmes de compression de données se sont avérés un analogue pratique de la définition de la complexité des groupements selon la TIA, et on analyse la possibilité d'utiliser un tel algorithme pour produire une mesure de la complexité visuelle.

Des études de la recherche visuelle reposant sur des textures visuelles et des groupements de formes visuelles ont démontré que certaines tâches de recherche sont indépendantes du nombre d'éléments ou de l'apparente complexité du groupement, tandis que d'autres varient en durée en fonction de la complexité du groupement. Il semble que certains des inputs sensoriels convergeant vers le système visuel soient assez faciles à distinguer les uns des autres (ce que l'on appelle le « pop-out » visuel), tandis que d'autres, guère plus compliqués en apparence que les inputs simples, sont difficiles à distinguer. Le concept de la séparabilité linéaire de l'arrière-plan de la recherche et des formes cibles explique en partie la différence : si les cibles et l'arrière-plan sont séparés par une partition linéaire dans l'espace de visualisation sensoriel, la tâche est facile; sinon, la tâche est difficile.

La convergence de concepts issus de la TIA, de la théorie de l'apprentissage perceptuel et de la théorie du circuit neuronal a mené à la synthèse suivante : selon la théorie de l'apprentissage perceptuel, une forme ou un groupement visuel simple génère de l'activité aux niveaux supérieurs du système visuel (au delà de l'aire 17 et jusqu'au cortex temporal médian), tandis qu'une forme ou un groupement complexe génère de l'activité depuis ces aires plus centrales jusqu'aux aires d'inputs sensoriels plus périphériques du système visuel. En général, les formes simples activent des aires visuelles centrales et les formes plus complexes activent des aires visuelles centrales et périphériques. Les formes simples ont des codes plus courts selon la TIA, ce qui signifie qu'elles activent des zones corticales moins nombreuses et moins spécifiques; les formes complexes ont des codes plus longs selon la TIA, ce qui signifie qu'elles activent des zones corticales plus nombreuses et plus spécifiques. Les théories du circuit neuronal convergent vers une explication de la perception visuelle comportant des couches multiples de circuits neuronaux qui se renforcent, chaque couche correspondant à un niveau supérieur de généralité. Les formes simples activent sélectivement des circuits d'ordre supérieur sans qu'il soit précisé quels circuits d'ordre inférieur sont actifs; les formes plus complexes activent des formes spécifiques à la fois d'ordre supérieur et d'ordre inférieur. On analyse l'application de ces principes et d'autres principes d'organisation de la visualisation dans l'étude de la performance basée sur la visualisation.

TABLE OF CONTENTS

INTRODUCTION 1

Vision and Visual Perception 1
Gestalt Psychology 1
Organization of the Review 2

GOOD FORM AND THE MINIMUM PRINCIPLE 3

How to Change Memory for a Visual Form 4
The Form as a Member of a Set 5
Information Theory and Form Discrimination 5
Information and Form Complexity 7
Dimensional Redundancy and Relevance 8
Information Separability and Integrality 10
Summary: Complexity and Distributional Constraint 10

CODING THE SINGLE VISUAL FORM 11

General Coding Rules 14
Summary: Coding the Individual Form 15

ALGORITHMIC INFORMATION THEORY 15

Simplicity and Probability 16
Complexity and Data Compression 18
Randomness, Objective Complexity and Subjective Complexity 18
Summary: Algorithmic Information Theory 19

VISUAL ARRAYS 19

Parallel Processing 20
Textons and Features 20
What Kind of Search? 21
Conspicuity and Discrimination 22
Eye Movements and Conspicuity 23
The Hick-Hyman Law 24
Summary: Array and Form Complexity 25

PERCEPTUAL LEARNING AND NEURAL CIRCUITS 26

Neural Circuit Theory 27
Colour and Linear Separability 29
A Synthesis of AIT, Perceptual Learning Theory and Neural Circuits 30

COMPLEXITY OF REAL DISPLAYS 30

SUMMARY 31

REFERENCES 33

LIST OF FIGURES

FIGURE 1. UNCONSTRAINED RANDOM POLYGONS (MAVRIDES AND BROWN, 1969).....6
FIGURE 2. CONSTRAINED RANDOM POLYGONS (MAVRIDES AND BROWN, 1969).....7
FIGURE 3. FORM DIMENSIONS. SHAPE, GAP ORIENTATION, LINE TYPE AND LINE ORIENTATION DISCRIMINATE EACH FORM FROM EVERY OTHER, BUT DIMENSIONS WOULD BE REDUNDANT IN A TWO – CHOICE CLASSIFICATION TASK. FROM GARNER, 1970. 9
FIGURE 4. AMBIGUOUS FIGURES THAT WILL BE SEEN AS TWO-DIMENSIONAL AS THEIR TWO-DIMENSIONAL COMPLEXITY DECREASES, WHICH IT DOES FROM LEFT TO RIGHT. FROM EXPERIMENTAL PSYCHOLOGY 3RD EDITION BY KLING © 1971. 12
FIGURE 5. STRUCTURAL INFORMATION LEVELS OF LINE FIGURES USING LEEUWENBERG'S CODING SYSTEM. FROM LEEUWENBERG, 1968.....14
FIGURE 6. GROSSBERG'S NEURAL CIRCUIT THEORY. INPUT (A) FIRES CELLS AND INHIBITS AROUSAL NEURONS (*a*). NEURAL CODE ACTIVATES NEXT LEVEL (B), WHICH RETURNS A MISMATCHED SIGNAL (C), RESULTING IN AROUSAL WHICH RESETS THE SYNAPTIC WEIGHTS, LEADING TO NEW MATCHING OF INPUT TO SECOND LEVEL. FROM GROSSBERG (1980). 27
FIGURE 7. A FEATURE DETECTOR (B) AND A HIERARCHY OF FEATURE DETECTORS (C) FROM RAO AND BALLARD, 1999. 32
FIGURE 8. EXAMPLE OF A COMPLEX DISPLAY FROM BURNS (2000) WHICH PRODUCED THE BEST FAULT DIAGNOSIS PERFORMANCE.....32

Introduction

Vision and Visual Perception

Vision is the most complex sensory system (Kandel, Schwartz & Jessell, 2000). It represents a physical world whose external stability and motion are maintained despite changes in visual input caused by movement of the perceiver (MacKay, 1962). It represents spatial dimensions with multiple mechanisms including motion parallax, aerial perspective and relative position for long distances and by calculating from binocular image displacement for short distances. It represents differences in surface composition and reflectance by differences in apparent brightness and colour. It identifies and tracks objects as they move, and it guides our movement through space and time. And because vision includes created images like letters and graphics, it is a medium for human culture.

In the nineteenth century, psychologists separated themselves from philosophers by claiming scientific credentials, and then they set to work to develop a scientific understanding of vision. They were guided by two older philosophical approaches: nativism and empiricism. Nativism emphasized the inherited competencies of the perceiver to analyze and understand visual input, while empiricism assigned the explanation for some aspects of visual perception to interactions between the perceiver and the environment.[1] Early perceptual theorists placed different emphases on analytical or synthetic approaches to the data. The effort to isolate and analyze elements of perceptual experience led to studies like "On the production of an artificial hiss" (Titchener, 1914), and "A qualitative analysis of tickling: its relation to cutaneous and organic sensation" (Murray, 1908). G.T. Fechner (1860/1966), an early analyst, quantified the relationship between an isolated part of the sensory world – for example, the physical intensity of a light source measured in lumens – and the resulting psychological experience of brightness. The tools of modern psychophysics and psychometrics like magnitude estimation scaling (Stevens, 1975) are direct descendants of and improvements on Fechner's analytical idea that elements of perceptual experience can be measured and then related to measurable components of the physical (and now the social) world.

Gestalt Psychology

While the analysts hoped to explain perception by first understanding its component parts like hisses, tickles and brightness and then developing the rules that joined the parts together in conscious experience, the synthesists went to the laboratory and produced visual experiences that the analysts could not explain. Gestalt psychology (Koffka, 1935; Kohler, 1947) was the result. Apparent motion (the phi phenomenon) is observed in a dark room between two spatially separated lights that are turned on and off in alternation. When the timing and distance parameters are right the apparent motion is indistinguishable from the real motion of one light moving back and forth in a dark room.

---

[1] Boring (1959), Boring (1942) or Murphy (1949) can be consulted on the evolution of theories of visual perception.

The phi phenomenon is the founding illusion of gestalt psychology (Hochberg, 1957; Rock and Palmer, 1990).

The gestalt law of prägnanz, as stated by Koffka (p. 110) is that "psychological organization will always be as 'good' as the prevailing conditions allow." This was rephrased by Hochberg (1957) as the "principle of 'simplicity' or 'maximum homogeneity.' " A gestalt good organization is a simple organization. The apparent motion of one object is a simpler perceptual organization than the alternate appearance and disappearance of two objects. The law of prägnanz implies that although the prevailing conditions (the sensory input) put a limit on the goodness, or simplicity, of a percept, every percept will be as good or as simple as the input allows. Gestalt psychology stated the problem of defining the connection between sensory input and perceptual simplicity or complexity, and so motivated much of the research reported in this review.

According to Hochberg, another outcome of the gestalt revolution was that "attention should be focused on higher-order variables of stimulation." This meant that perception could not be explained by sensory "atoms" held together by associationistic laws. Instead, the sensory stimulus should be described to account for its internal relations and its variation over space and time. Describing the higher-order invariants of the visual stimulus was a major accomplishment of J. J. Gibson, who analyzed the visual expansion pattern that pilots use to land airplanes (Gibson, 1955a, 1955b), and then extended his work on the analysis of higher-order perceptual variables in two influential books: *The perception of the visual world* (Gibson, 1950) and *The senses considered as perceptual systems* (Gibson, 1966). Gibson analyzed the patterns of stimulation that lead to perception of texture and texture gradients; he analyzed the variations in light and shade that indicate distance and form, and he emphasized the interactive aspect of visual perception which allows the perceiver to actively seek visual information from the higher-order invariants of the "visual world."

Organization of the Review

Thus the gestalt revolution motivated two parallel and complementary research efforts: one that sought to provide substance to the principle of simplicity or maximum homogeneity (often known as the minimum principle), and the other that sought to provide evidence for the role of higher-order variables in perception. Both efforts have contributed to understanding visual display complexity, and both will be reviewed here.

The first area to be reviewed is research on the perception of the single visual form. Two main approaches have been taken in using the minimum principle to understand the perception of a single form. One approach is to consider each form as part of a larger set, either implicit or explicit, and to derive properties of a single form from properties of the set. The second approach is to develop a description or code for a single form that can be interpreted as a measure of simplicity or complexity. The close relationship between these two approaches will be discussed.

A recent approach to the measurement of visual complexity called algorithmic information theory (AIT) attempts to reconcile the competing claims of modern gestaltists and modern associationists, who, like the old analysts, give a larger place in perceptual theory to individual differences in perceptual experience. Like the successful parts of gestalt theory (the unsuccessful physiological parts were abandoned a long time ago: see Sperry, 1958; Rock and Palmer, 1990), algorithmic information theory is computational and does not invoke knowledge about the central nervous system.

In the visual display context, research on higher-order stimulus variables means research on the textures and features of arrays. Much of this research asks how quickly and accurately people can identify particular features or textures, how well they can discriminate among different ones and how well they can search a display to identify a particular display element, usually called a target, against a background of other display elements that form a feature or texture background. This leads to consideration of the general problem of visual search in terms of target complexity and display complexity, and raises the related question of the role of attention in the perception of visual displays.

Hochberg's advice to search for higher-order perceptual variables came just a few years before Hubel and Weisel (1968) reported "edge detector" and "complex" cells – cells that defined higher-order perceptual variables – in the primate visual cortex. Advances in neural circuit theory, coupled with a better theoretical understanding of the mechanisms of visual attention and perceptual learning, make it reasonable to consider the role of perceptual learning and neural circuit theory in display complexity.

A discussion of real visual displays is followed by a summary of conclusions drawn from the experimental and the theoretical work.

Good Form and the Minimum Principle

"Good form" is the central idea that survived from gestalt psychology into the twenty-first century, but there is little left of its original content, as expressed, for example by Koffka (1935) or Kohler (1947). Kohler thought that the brain tissue in which visual input was initially represented acted on the trace of that input to dynamically modify the resulting perception towards an ideal or optimum representation (Kling & Riggs, 1971). The corollary was that the dynamic influences on the trace in the absence of sensory input could be studied as the trace changed with time. According to Koffka (pps. 493 – 495), the perception resulting from a visual input could change in several ways. If it was similar to a more regular percept it would be assimilated to that percept (*normalizing*); if it had some distinct feature which differentiated it from another percept the difference between the two would be exaggerated or sharpened (*pointing)* and if there were stresses brought about by its internal dynamics there would be *autonomous* changes. This is ambiguous: how does one know whether a trace will be normalized towards a regular percept, or pointed away from it?

Gibson (1929) reported experiments that he thought demonstrated autonomous trace changes. Gibson's experiments involved repeated reproductions of the same forms, and it

turned out that the changes he observed were associated with the re-exposure of the observer to the distorted copies of his own reproductions. A series of better-controlled studies showed that in the absence of any biasing effect of repeated exposure there is no systematic distortion in visual memory over periods of days or even weeks (Rock and Englestein, 1959; Hebb & Foord, 1954; Hanawalt, 1952). Whatever good form is, it does not predict changes in the perception or memory of an *isolated* single form.

How to Change Memory for a Visual Form

Despite evidence that there is no *autonomous* change in visual form memory, it is easy to produce a *predictable* change in form memory. Carmichael, Hogan and Walter (1932) showed that a name associated with the first appearance of a form would influence the way the form was later reproduced. An example is two circles joined by a straight horizontal line. When called "eyeglasses" this form was reproduced later as two circles joined by a curved line (the nosepiece of the eyeglasses). When called "dumbbells" the form was reproduced as two circles joined by a double line, the bar of the dumbbell. The obvious explanation for this is that the memory image associated with the word was associated with the memory image of the visual stimulus, and the reproduction resulted from a combination of the properties of the image associated with the word and with the visual input.

W.R. Garner (1962, p. 342) proposed that the memory for a visual form would change when the change made it easier to categorize the form in a concept formation or visual discrimination task. His example was that if a circle, square, triangle and trapezoid were to be discriminated and a small gap was placed at the bottom of each figure, the gap might be reduced or disappear in memory because it was not essential to discriminate one figure from the other. On the other hand, if four squares were to be discriminated, each one differing from the other only by the location of a small gap on one of the sides, that the gap would be maintained or even increased in memory. In Garner's words, "sharpening will occur for those aspects of the figure which are differentiating, and closure, or perhaps levelling, will occur for those aspects which are non-differentiating." Garner also went on to suggest that the differences between levels of a particular feature that discriminated a set of visual objects might be increased in memory.

Donderi, Case, Rossoff & Willis (1970) and Donderi (1973, 1990) demonstrated Garner's conjecture. They presented a visual discrimination task based on one dimension of a set of stimuli that differed on two dimensions, for example circle size and lightness, circle size and fill density, or ellipse orientation and size, and compared changes over time on the relevant versus the irrelevant dimension. In all three sets of experiments, the remembered differences on the relevant dimension increased relative to the differences on the irrelevant dimension. The differences were greater immediately after learning, decreased after a day, and increased again at delays of one or two weeks. Tests at each delay used different observers to avoid contamination of the results by previous testing of the same observers.

The Form as a Member of a Set

Emphasis on the visual stimulus as a member of an ensemble was inspired by the principles explained by Shannon and Weaver (1949) in *The mathematical theory of communication*. The basic idea of Shannon's information theory is that an improbable symbol in a message conveys more information than a probable symbol. If all of the symbols in every possible message were always the same, then the messages convey no information, because the receiver of the message knows no more after receiving the message than he did before it arrived. He could predict the message in advance with probability 1. But if two different messages might be received with equal probability, then receipt of one message reduces the *uncertainty* about which message occurred. Uncertainty reduced is information gained, and the amount of information contributed by knowing which of two equally probable messages was sent is one *bit* of information. Shannon defined the information value H of an individual symbol in a message consisting of a sequence of symbols as $H = -\log_2 p$, where p was the probability of the symbol occurring in the set of symbols from which the message was drawn.[2] From this basic definition he derived definitions and identities and proved theorems about the amount of information in an entire message, about the effects of transmission errors on the probability of coding and transmitting a message accurately, and about the relationships between messages sent and received in a communications system.

Early in his book Shannon introduced examples of alphabetic messages consisting of the 26 letters plus a space. The greatest amount of information conveyed by an alphabetic message occurs when successive letters are transmitted with equal probability and independently of every other letter. The result is a string of letters .... FFJUJ ZLPW ... that carries the maximum amount of information but bears no resemblance to language. But a message like "IN NO IST LAT WHEY CRATICT.." looks more like English. In this message, each letter is chosen so that the relative probability of occurrence of successive letter triplets, e.g. IST, SAT, WHE, matches the probability of these triplets occurring in English. This means that each letter is to some degree predictable from the two preceding it in the message. The result is that the message contains less *information*, but has more *structure*. In information-theory terms, the message is *redundant*, which is another way of saying that some parts of the message can be predicted from other parts of the message.

Information Theory and Form Discrimination

Information theory was immediately adapted to psychological problems (Quastler, 1954; Miller, 1956; Attneave, 1959), and has played a major role in the study of visual perception for the past half-century. In order to use information concepts in visual perception, it was necessary to make stimuli whose information content could be measured. This meant composing forms from perceptual subunits (symbols) whose probability could be varied over the ensemble of forms (messages). Fitts et al. (1956) developed stimuli of this type. They created 4 x 4 or 8 x 8 matrices in which each column

[2] Shannon used symbol H to recognize the contribution of R. Hartley, who first proposed that a logarithmic unit was an appropriate measure of information transmission.

of the matrix defined a different stimulus element. Numbers from 1 to 4 or 1 to 8 were chosen randomly to assign a height to each column, creating a "histogram figure." They introduced various forms of non – randomness across the set of figures by using mirror images to duplicate columns in some figures, and copying heights of some columns into other columns. Their performance measure was the speed with which a test figure could be identified among the rest of the figure set that was presented in several repeated rows of figures below the test figure. Random figures from sets created with no constraints, and therefore conveying the greatest amount of information per figure, were most rapidly discriminated from the test figures comprising the rest of the set. By analogy to Shannon's use of communication theory, the set of figures was an ensemble of messages, each message was more or less distinguishable from the other messages by the amount of independent information it conveyed, and the more distinguishable it was from the rest of the figures on the basis of its information content, the faster it could be distinguished from the rest of the messages when they were presented together.

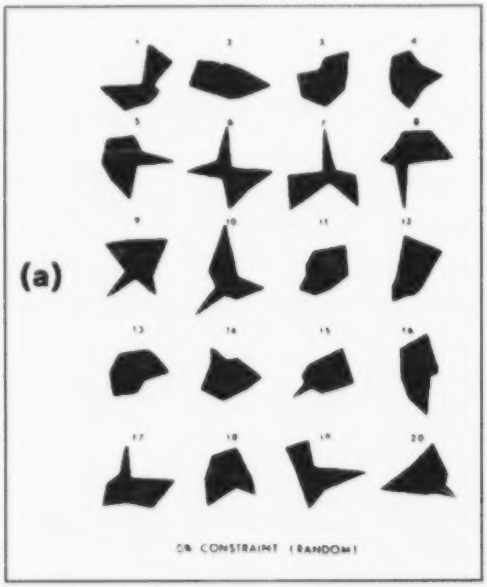


Figure 1. Unconstrained random polygons. From Mavrides and Brown (1969).

(Reprinted with permission. Copyright Psychonomic Society, 1969.)

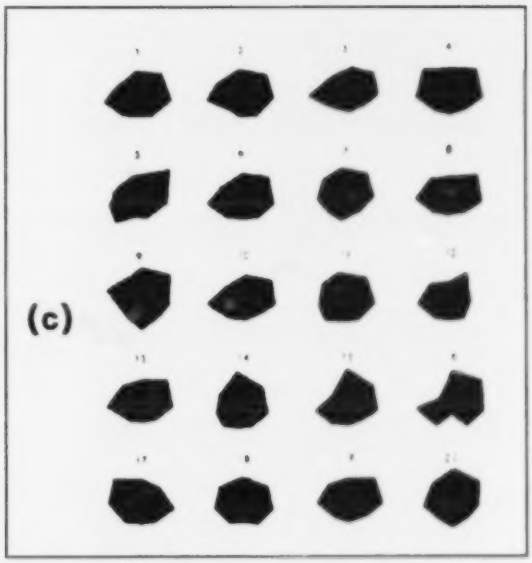


Figure 2. Constrained random polygons. From Mavrides and Brown (1969).

(Reprinted with permission. Copyright Psychonomic Society, 1969.)

Information and Form Complexity

Anderson and Leonard (1958) used three sets of eight – column histograms that differed in the degree of constraint (redundancy). There were eight figures per set. Observers were tested on the standard recognition task (turning over copies of figure in successive rows of figures), on a paired - associates learning task with the figures as stimuli, and on a reconstruction task where an adjustable model was used to copy the figure on successive 0.5 sec exposures of the original. They found that the greater the degree of randomness in the figure set, the faster the figures could be recognized and the better they were as stimuli in a paired – associates learning task. But the constrained forms, with less randomness, took fewer short exposure trials to reproduce accurately, although the random figures were reproduced more accurately on the first reproduction trial. Generalizing from this result, forms with more information (more "complex" forms) are easier to discriminate, while forms with less information ("simpler" forms) are harder to discriminate on the first reproduction trial, but take fewer trials to accurately reproduce.

Mavrides and Brown (1969) varied the Fitts et al. technique by transforming the histograms into polygons. They substituted the length of eight equally spaced radii for the eight column lengths of the histogram. They reproduced three sets of polygons from histogram data with different degrees of constraint: zero constraint (random polygons), 40% and 70% constraint, constraint being defined by limitations on the selection of radius lengths for the figures. The constrained sets are more regular. Figure 1 shows a random set, and Figure 2 a 70% constrained set. The constrained polygon sets produced

more errors on a single reproduction trial where the polygon was exposed briefly and then had to be reproduced, in part, on a test sheet. They found the same results for histograms constructed by the same rules. Another experiment in which observers found a matching polygon among a set presented along with the test polygon yielded similar results. This evidence, like the experiments of Fitts et al. with histograms, shows that constraints within a set which reduced the information (relative to the rest of the set) conveyed by each stimulus, made the forms harder to reproduce accurately on a first trial, and harder to discriminate from the rest of the set. Unlike Anderson and Leonard, they did not test the number of exposure trials required to accurately reproduce the forms.

Mavrides and Brown carried out another type of analysis that suggested an entirely different approach to the analysis of visual forms. On the basis of extensive factor analyses of form complexity measures, they isolated seven physical measures: 1) variance in interior angles, 2) mean of the X coordinates, 3) mean of the Y coordinates (zero for the centre of the form), 4) second moment of the deviation of side length from that of a regular polygon of the same area, 5) variance of the side lengths, 6) largest radial length, 7) perimeter length. Form reproduction performance was predicted independently of the constraint generation conditions by multiple regression equations including each of the seven measures for every form. Thus the ensemble constraint conditions producing differences in the information content of each form also produce physical differences in the description of single forms.

Earlier, Arnoult (1960) had shown that observer's subjective complexity measures for individual random polygon forms could be very accurately predicted by a combination of measures of the number of independent sides, the ration of perimeter^2/area, and a measure of form symmetry. Thomas (1968) also measured the complexity of a set of 20 random polygons in the process of carrying out multidimensional scaling of the similarity ratings of each of the 190 polygon pairs. A best-fitting two-dimensional similarity solution included one dimension labelled "complexity" which was entirely accounted for by the number of angles in the polygon. Deriving measures for a single form from properties of a set of forms is a topic that will recur in the later discussion of algorithmic probability measures of form complexity.

Dimensional Redundancy and Relevance

Imposing constraints on form dimensions changes the information content of each form independent of the tasks to be carried out with it. Once a task is imposed – say placing each form into one of two categories – then another aspect is added to the description of each dimension of each form: its task relevancy or irrelevancy. For example, if forms are to be classified only on the basis of size, then size is relevant and shape is irrelevant. If forms are classified only on the basis of shape, then shape is relevant and size is irrelevant. Classification tasks can be arranged to make the same dimensions sometimes relevant and sometimes irrelevant. Figure 3, from Garner (1970), shows a set of sixteen forms whose dimensions would be non-redundant for a sixteen – item paired – associates task, but redundant for a two – choice classification task.

Bourne and Haygood (1959) independently varied the number of redundant relevant dimensions and non-redundant irrelevant dimensions to determine their effects on the number of errors to criterion, and number of trials to criterion, on a two-choice concept identification task. Performance was better with an increase in redundant relevant dimensions, was worse with an increase in the number of non-redundant irrelevant dimensions, and the effect of the number of irrelevant dimensions (1, 3 or 5) was greater with only 1 relevant, rather than more relevant dimensions. In a second experiment, the number of redundant irrelevant dimensions was varied from 1 to 5, for two different conditions of number of non-redundant relevant dimensions. There was an increase in the number of errors to criterion with an increase in the number of redundant irrelevant dimensions, but the increase was roughly the same for both the generally harder 4-choice (2 relevant dimensions) and the easier 2-choice (1 relevant dimension) task.

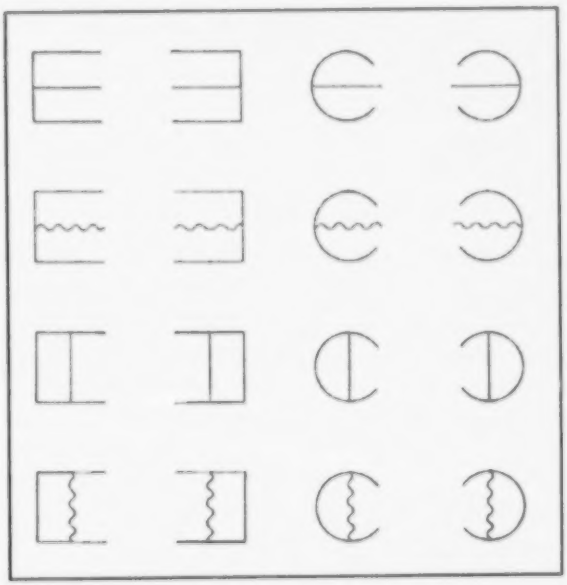


Figure 3. **Form dimensions. Shape, gap orientation, line type and line orientation discriminate each form from every other, but dimensions would be redundant in a two – choice classification task. From Garner (1970).**

(Reprinted with permission of *American Scientist*, magazine of Sigma Xi, The Scientific Research Society)

These experiments show that the value of information measurement is contingent on what the information in a visual form will be used for. The task contributes to understanding the role of the particular information in a visual form.

Information Separability and Integrality

For information measurement to adequately describe a visual form in the same way it is used to describe the communication of messages, the form must be composed of equally discriminable parts, like the symbols in a message. But aside perhaps from random histograms and polygons, visual objects are not like that. Even if it is possible to separate different components of a visual object like size, shape, brightness and colour, the blend of these components into a single object changes their characteristics. Lockead (1966) and Garner (1974, pps. 102, 110 - 115) introduced the concepts of *integrality* and *separability* of dimensions. However the different basic sensory qualities are distributed across a set of visual objects, visual characteristics like hue and brightness combine to form a perceptual experience in which it is difficult to analyze out the original components. They produce an *integrated* perceptual experience. By contrast, visual characteristics like size and shape remain conceptually separate, or *separable*, to an observer and do not form an integrated perception. So the distribution of hue and brightness, and size and shape, can be varied independently across a set of visual objects to change their informational characteristics, but this does not equate their perceptual relevance.

Garner showed that when observers judged the similarity of visual objects that differed on integral dimensions, they combined the differences between the objects on each integral dimension using a Euclidean spatial model, but when they made similarity judgments to objects that differed only on separable dimensions, they used a "city-block" metric. The aspect of "similarity" of combined perceptual experience was used more often by observers who were describing or classifying visual objects containing integral dimensions, and the aspect of "classification" of separable dimensions like size and shape was used more often by observers who were describing or classifying visual objects based on separable dimensions.

Summary: Complexity and Distributional Constraint

Constraint on the random selection of components of a visual object produces objects that are harder to discriminate but easier to reproduce. Object complexity is produced by the absence of constraint in selecting object components; object simplicity is produced by imposing constraints. *Integral* object components like hue and brightness fuse perceptually to give a unitary perceptual experience, while *separable* components like size and shape can be more easily analyzed out of the visual object. Integral and separable components scale differently when object similarity is measured, but the probabilities of both integral and separable components across the object set can be controlled in the same way. Local (non-distributional) physical properties of individual visual objects like random histograms can be measured, and such properties like the number of angles, the perimeter/area ratio, and moments of the distribution of inflection points on the perimeter combine to provide good predictions of the subjective experience of "complexity." Thus there is a direct influence on local object properties of the rules

that govern the generation of an ensemble of objects, and thus a direct influence of these rules on subjective complexity as predicted by these local properties.

Coding the Single Visual Form

Mavrides and Brown (1969) showed that subjective complexity judgments for random polygons were determined by physical measures that could be calculated for each polygon, and that the generating rules (random or constrained) that determined the information content of the set of polygons directly influenced the physical measures for a particular polygon. This suggests that the properties of a single visual form, taken in isolation, can be derived from the generating rules that produced it, and by implication, can be related to the other visual forms generated by the same rules. Garner (1970, 1974) showed that this was true by demonstrating that the subjective "goodness" of a single visual form could be predicted by calculating the size of the set which could be generated from that visual form by rotating or reflecting it. The smaller the size of the alternative stimulus set the "better" the figure. He tested this idea with patterns composed of five dots placed within the cells of a 3 x 3 matrix. Each pattern, when reflected or rotated within the matrix, produced a larger or smaller set of other patterns (reflection or rotation around an axis of symmetry would produce an identical pattern, and therefore not increase the total number of patterns). Patterns that produced few other patterns on reflection or rotation were rated as being "better" figures than those that produced many other patterns. Garner showed, in other words, that pattern generating rules work both ways to determine the goodness, or simplicity, of a visual form. Constraints (redundancies) imposed by the rules that generate visual forms produce sets of forms that are judged to be simpler, that are harder to discriminate from one another, and are easier to reproduce. Visual forms that are judged as simpler are forms which, when subjected to transformation rules like reflection and rotation, produce fewer other forms.

Complexity rules like those produced by Arnoult (1960) and Mavrides and Brown (1969) are related to Garner's transformation process because forms with greater complexity as measured by rules that count angles, or the ratio of perimeter to area, will produce a larger set of alternative forms when rotation and reflection transformation rules are applied to them. This suggests that there may be a convergence between complexity rules that are intrinsic to a particular form and generation rules that produce a given set of forms. Generation rules that do not impose constraints on the construction of visual forms guarantee that a larger set of forms can be produced, and that any one form (including a simple one) will have a lower probability of occurrence in the set. Lower - probability forms transmit more information. Generation rules that impose constraints limit the number of forms that can be produced and so guarantee that there will be less variation within and between forms, and since these rules reduce the number of forms that can be produced, each form will have a higher probability of occurrence in the set. Higher – probability forms transmit less information. In short, a less probable form is a more informative (complex) form, and a more probable form is a less informative (more probable) form.

Hochberg (1957) concluded that, with respect to the residual of the gestalt revolution, "the principle of 'simplicity' or 'maximum homogeneity' has proved the most general." By this he meant that the perception of a visual form will tend towards the simplest possible description of that form, in some descriptive language or other. Hochberg and Brooks (1960) had observers rate a large set of ambiguous line drawings (that could be seen either as two – dimensional or three – dimensional objects) for whether they appeared to be more two-dimensional or three-dimensional. They then found physical measures of two-dimensional complexity that are very much like those found by Arnoult and Mavrides and Brown: number of angles, number of line segments, and variability of the angles. The greater the two-dimensional complexity, the higher the three-dimensionality rating of the figure (see Figure 4). They confirmed this analysis by obtaining three-dimensionality ratings for a new set of ambiguous figures, and showed that the same two-dimensional complexity measures predicted three-dimensionality of the new figures equally well.

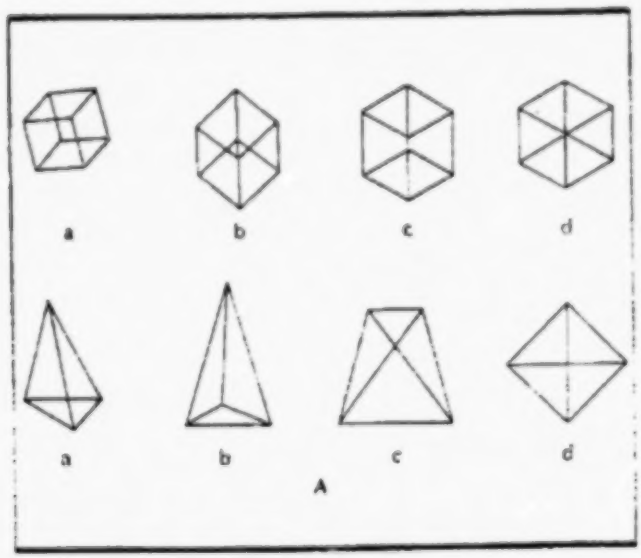


Figure 4. **Ambiguous figures that will be seen as two-dimensional as their two-dimensional complexity decreases, which it does from left to right. From Experimental Psychology 3rd edition by Kling © 1971.** (Reprinted with permission of Wadsworth, a division of Thomson Learning: www.thomsonrights.com Fax 800 730-2215)

The complexity rules like those used by Hochberg, Garner, Arnoult or Mavrides and Brown apply to line drawings, random histograms, random polygons or dot patterns with a limited range of visual components. Another set of complexity measurement principles

for visual forms was developed by Leeuwenberg (1968). Leeuwenberg's goal was to describe the complexity of visual patterns by assigning each pattern a code whose length was a measure of the pattern complexity. He started with three very general propositions. The first was that the perceptual encoding of any pattern was based on an irreducible pattern unit, called a "non-complex property". The second was that a pattern code was built on successive earlier steps, tracing back eventually to one or more "non-complex" properties of the pattern. The third was that the code generated a single perceptual object from the pattern but excluded other perceptual objects that could be constructed from the same pattern, as in the case of an ambiguous figure.

He detailed these general propositions with a series of operators and rules that applied to two-dimensional visual patterns. The rules prescribed how successive operators should be applied in order to derive an object description from the non-complex properties of a particular pattern. He described the application of these rules and operations in the following sentence, italicized in the original: "*We shall therefore equate the information of a figure to the number of operations – corresponding to independent and indivisible aspects – which must be successively applied in order to rob the given figure of its content*". (Leeuwenberg, 1968, p. 10). The operators included actions called reduction, differentiation and reversal. Reduction meant the process of reversing a pattern property until it became a "non-complex property;" therefore, the number of reductions was a measure of how far a pattern deviated from a "non-complex property" which represented the minimum amount of complexity. Differentiation meant taking the difference between adjacent pattern elements which, if constructed according to some repeating rule, thus reverted to a single pattern element of a constant type. And reversal meant identifying pattern elements of the form (b, a) as a reversal of the basic elements (a, b), and thus expressing them as a simple modification of the basic elements (a, b).

Leeuwenberg reported a variety of studies that related cognitive and perceptual performance measures to his information measure. He had observers rate the complexity of a set of 24 continuous figures and another set of 24 dot figures. All of the figures were visible at the same time, and his instructions emphasized that the observers should carefully analyze each figure, while keeping all of the figures visible. The figures were sorted into twelve groups of increasing complexity, but observers were free to sort as many figures into each group as they wished. For the continuous figures, the average complexity ratings correlated 0.97 with the structural information measures (see Figure 5). For the dot figures, the correlation was 0.83. Leeuwenberg than analyzed the discrepancies between the information structural information in each figure and the figure's judged complexity. The discrepancies are quite interesting for the dot figures. He found that complex dot figures that were constructed from repeated "simpler" elements (elements having internal "good form") were judged lower in complexity than figures that had the same Leeuwenberg measure of complexity, but a less regular internal structure. Since Garner (1962, 1970) demonstrated that dot figures with the same kind of "simpler" elements were rated as less complex, and attributed this to the smaller subset of forms that could be produced by rotations and reflections in the plane of a good figure, it appears that Garner's principle needs to be added to Leeuwenberg's principle to better account for Leeuwenberg's results.

General Coding Rules

Leeuwenberg (1969) then developed a concise notation for the codes that he developed in his 1968 thesis, and used it to re-evaluate a series completion task that had been studied earlier by Simon and Kotovsky (1963). Simon (1972) then comprehensively reviewed the development of symbolic codes for perceptual complexity, including the work of Leeuwenberg (1969), his own work (Simon and Kotovsky, 1963), and others. He noted that all of the coding theories represent the sequential pattern of elements in a visual form, and in all of the coding theories the length of the representation (the code length) is the measure of form complexity. The coding languages use all or some of a set of operators that describe the repetition of pattern elements, the reversal of pattern elements, the identity of elements, the succession of elements in defined alphabets, recursive operations on sequences of symbols, working memories (where some code element can act as a pointer to a member of a remembered alphabet), and push-down and pop-up operations that might apply to code lists held in short-term memory. Simon cited data from several papers, and generated some more of his own, which showed that the code lengths, and therefore the complexity measures, from several of these theories were highly correlated with each other and with performance measures generated with the patterns. He concluded that there is no single coding language for patterns that is *the* best language. Instead, "we need to determine the conditions (sociological and psychological) under which subjects will adopt one or another coding scheme in handling patterned sequences."

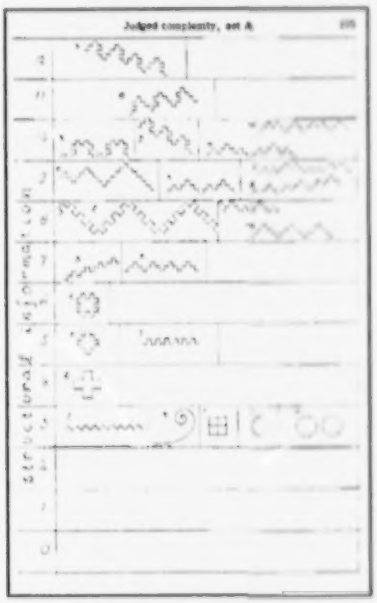


Figure 5. **Structural information levels of line figures using Leeuwenberg's coding system. From Leeuwenberg (1968).** (Reprinted under Fair Usage Policy due to year of publication – publisher unable to find copyright holder.)

Summary: Coding the Individual Form

The research on the single form reviewed so far can be summarized as follows: As a result of the influence of information theory on perception psychology, the gestalt idea of " good form " was converted into the study of the distributional properties of the components of a visual form over a set of similar forms. The distributional properties predicted both subjective judgments about the forms and performance using the forms as stimuli. Then it was observed that complexity measures calculated for individual forms without reference to the distributional properties correlated well with perceptual judgments of good form and with perceptual performance based on the forms. Then it was observed that the simplicity or goodness of a particular form was related to the size of a set of forms that could be generated by rotating or reflecting the target form in two dimensions. Thus it appears that the form goodness is a property of the form that depends on the rules used to generate the form, that goodness can be calculated by direct measurement on the form, and that it can equally well be determined by the size of a set of forms that can be generated from the particular form by a specified set of transformations (rotation and reflection). The relationship is easiest to understand in the context of random polygons. A constrained generation rule will produce simpler polygons that are also more similar to each other, and each polygon so created will generate lower measures of complexity based on things like variation in the number and variability of angles or sides. In turn, a simpler polygon will generate a smaller set of different polygons when rotated or reflected in the plane. The appearance of the single polygon is influenced by, and in turn reveals, the set of potential other polygons with which it could have been generated.

Algorithmic Information Theory

The literature cited shows that there is a connection between the distributional rules used to produce a set of forms and the complexity code for a single form. The theoretical link between distributional rules and form complexity was made with algorithmic information theory (AIT, Solomonoff, 1986). AIT provides a way to measure the probability of a string of symbols and then to link that probability to the complexity of the symbol string. In order to apply AIT to visual form perception, the form has to be translated into a string of symbols by scanning it into a computer and identifying the form with the file of bitmap pixels that reproduces the form.

In AIT, the definition of algorithmic probability of a string, s, is "the probability of that particular string being produced as the output of a reference universal Turing machine with random input. It is approximately 2^{-p} where p is the length of the shortest program, p, that will produce s as an output. p is the Kolmogorov complexity of s, one form of Algorithmic Complexity" (Solomonoff, 1986, p. 2).

Thus the probability of a string representing a visual form is the length of the shortest program which when applied to a Universal Turing Machine, will generate the string s that is identified with that form as an output. Since the Turing machine has a random input, the probability of any particular string being produced by random input to the

Turing machine decreases with increasing length of the output string. Clearly a Turing machine (which produces only 1s and 0s as output) will produce a random 1 or 0 very soon after it begins to operate on a random "program tape" of 1s and 0s. If a visual form is all one colour and brightness over an entire large field (like a blank computer screen) then a short program tape of 1s and 0s will soon produce the short string encoding that form, and the form complexity will be low.

The probability that the machine will produce a particular sting of 1s and 0s from a "program tape" of length p decreases with increasing length of the required output string, because only one "program tape" of that length will encode the output string, and there are 2^{-p} possible tapes of length p. So AIT postulates that the complexity of the output string representing a visual form is, as suggested by Leeuwenberg (1968, 1969) and Simon (1969), related to the length of the code required to describe the form.

Solomonoff proposed that the algorithmic probability of a string is the sum of the probabilities of all the possible "program tapes" that could generate that string when used as input to a Turing machine. If a particular short tape produces the string, there is no reason why a longer tape might not also produce the same string. A limiting argument can even explain why, given this definition, simpler shapes produce fewer other shapes on reflection and rotation (Garner, 1970, 1974). If we assume for practical purposes that there is a finite limit to the length (and therefore the number) of program tapes that can be fed to the Turing machine, then the larger the number of those tapes that produce the same output string (e.g. the visual form under consideration), then the smaller is the number of other visual forms that can be produced by any program tape. Tapes corresponding to reflections and rotations of the target form will, if the form is simple enough, simply reproduce the target form. Thus a higher probability of producing the form in question – the definition of a simpler form – means a smaller probability of producing any other form – thus the limitation on the set size of forms produced as alternatives to the first form.

Simplicity and Probability

AIT connects the concepts of "simplicity" and "probability" directly, and allows a connection to be made between two principles that have long dominated thinking about the organization of visual perception. One principle is the "likelihood principle," which is that a given visual sensory input will lead to the perception of the most likely distal object, given the input. This puts visual perception into a hypothesis-testing framework, where the "data" is the visual input, the "hypotheses" are the possible distal objects that the input represents, and visual perception can be formulated as a Bayesian probability decision:

$$p(H/D) = p(D/H) \cdot p(H) / p(D),$$

where H stands for perceptual hypothesis (the percept), D stands for the sensory input data, and the visual system fixes as the percept generated by sensory input D the perceptual hypothesis H that is maximized by the formula. In order to carry out this

calculation, the visual system needs probabilities for both H, D/H and D: in other words, it has to have a set of stored probabilities handy: the probability of the current perceptual input D, the probabilities of possible perceptual hypotheses H, and the probabilities of the data D given each hypothesis H.

The "simplicity" approach to visual perception assumes that the visual system chooses a perception based on an information-theory analogue to Bayes' theorem:

$I(H/D) = I(H) + I(D/H) - I(D)$.

In words: the information (complexity) of a possible percept, given the sensory input D, equals the complexity of the percept plus the conditional complexity of the data D given the percept H, minus the complexity of the data D. The perceptual hypothesis chosen in any situation is the one that minimizes this information function. In order to solve this equation, the perceptual system has to store complexities for three terms: the perceptual hypothesis itself, the sensory data, and the complexity of the sensory data given the hypothesis (van der Helm, 2000, p. 781).

This can be illustrated concretely with an example from the Hochberg and Brooks study of the conditions for judging an ambiguous line drawing to be either two-or three-dimensional (See Figure 4). The alternative hypotheses are a two-dimensional or three-dimensional perceptual object given a fixed sensory input. The two-dimensional information hypothesis for ambiguous figure a is longer (contains more descriptive elements) than the information hypothesis for figure d, which because it is more simply symmetrical, contains fewer elements. The three-dimensional information code for figure a and figure d is the same: the code for a cube with a single parameter: edge length. The two-dimensional solution for $I(H_{2d}/D)$ will be larger for figure a than for Figure d, while the three dimensional solution $I(H_{3d}/D)$ is the same for both a and d. A is therefore more likely to be seen as three-dimensional because the three-dimensional solution minimizes the solution for a more completely than it does for d.

The connection between the likelihood principle and the simplicity principle lies with the identification, in AIT, of the complexity of a percept as p, which is the length of the string that generates the percept and also the (negative) exponent of the term that defines the Kolmogorov probability of the percept. Probability and complexity are directly linked: the most probable percept is the least complex, and the least probable percept is the most complex. The visual system can be said to minimize either the complexity of a percept, given the input, or to maximize the probability of the percept, given the input. In order to generate the *most probable* percept conditional on the sensory input it generates the *simplest* percept conditional on the sensory input (Chater, 1996; van der Helm, 2000)[3].

[3] van der Helm (2000) does not accept Chater's (1996) automatic identification of probabilistic and informational descriptions because he disagrees with certain formal probability definitions in the probability model. This does not is in my opinion negate the practical convergence between the two approaches.

Complexity and Data Compression

The entire enterprise of identifying the complexity or probability of a bit string with the length of the minimum program required as input to a universal Turing machine to reproduce that string is very much like the enterprise of compressing data to facilitate data transmission – an original goal of Shannon's information (communication) theory. In fact, a practical analogue of the universal Turing machine might be found in data compression algorithms. The coding algorithm of a compression program first generates a compressed file of length s that, when passed through the decoding algorithm, regenerates the original bit string. Kolmogorov complexity of the original bit string is p, which is defined as the length of the shortest program that can reproduce the string. If p is taken as the file length of the compressed file (program) that when passed through the decoding algorithm, reproduces the original bit string, then p might stand in for the Kolmogorov complexity of the original string. It was not necessary to generate $2^{-|p|}$ random programs of length p to reproduce the original string because compression algorithms are designed to do the job on one trial. The compression algorithm acts as an inverse Turing machine to generate the program p, and then as a Turing machine to regenerate the original bit string from the program p, just as intended in the definition of algorithmic probability.

According to Solomonoff's AIT invariance theorem (Chater, 1996, p. 573, van der Helm, 2000, p. 777), the length of the code s that is used to define the probability (or complexity) p of a symbol string is invariant, up to a multiplicative and additive constant, regardless of the language used to encode it. That means that if we consider file compression programs as alternative coding languages for a bit string, the compressed file lengths of the same bit strings expressed in different coding languages should be invariant up to a multiplicative and additive constant.

This leads to the possibility that data compression algorithms might provide a practical alternative to defining the complexity of a bit string – and therefore of a visual form encoded as a bit string – as the length of the generating program input to a Turing machine. The compressed file length might act as a stand-in for the Turing-based measure of algorithmic complexity. This possibility has been investigated and is reported in the contract report on the experiments.

Randomness, Objective Complexity and Subjective Complexity

A random string of symbols cannot be generated by a Turing machine using a program that is shorter than the string itself. That is because there is no regularity in the random symbol string that could be reflected in instructions coded in the program. Instead, the best possible program is a copy of the random string to be reproduced, with the instruction that every bit in the program is to be reproduced in the string. This means that by AIT rules, a random string is the most complex object of its length.

But random images may be judged to be less complex than images that are structured in complex, non-random ways. Or, as Wolfram (2002) suggested, the perceived complexity

of an image is a function of the most complex structure in the image that humans can detect. We cannot detect the structure in a random image. This suggests that the subjective complexity of an image will be at a maximum somewhere between simple order and complete randomness, and that complete order and complete randomness are not that perceptually different. In the extreme, an entirely black computer screen would be judged not much less complex than a screen generated by a random collection of red, green and blue pixels. Not only that, but human visual sensitivity decreases at high spatial frequencies, and very complex, high-frequency non-random patterns may not be detected because their details are below the threshold of visual discrimination. From this it follows that subjective complexity might be an inverted U-function, rather than a monotonic function, of algorithmic complexity. There is at present no objective evidence about this.

Summary: Algorithmic Information Theory

Algorithmic Information theory (AIT) provides a link between the probabilistic and the informational interpretations of form complexity. By describing a visual form as a bit string (a practical description in the computer era), AIT provides an abstract method for quantifying the complexity of the form. That measure is the length, p, of the "input" bit string that generates the "output" bit string corresponding to the visual form when the input bit string is fed into a Turing machine. Since this abstraction supposes the input to the Turing machine to be a random bit string, the probability of generating the particular bit string that produces the output big string is 2^{-p}, which decreases with increasing string length p.

While one does not feed random bits into an abstract Turing machine to generate bit strings in everyday life, the AIT measure has a practical analogue in data compression programs that do analyze the bit string corresponding, for example, to a visual display, and then produce a shorter bit string that when passed back through the data compression program reproduces the original string. The shorter string is the analogue of the input to the Turing machine that produces the display "output" bit string; its length, p, is the analogue of the AIT measure, p, of the complexity of the output bit string.

Visual Arrays

A practical question connected with visual form perception is what happens when many forms appear at the same time in an observers' visual field. They become part of a larger visual experience called a visual texture or a visual array (Gibson, 1950) The defining characteristic of a visual array is that individual forms lose some of their distinctiveness when they become parts of the array. The visual world is a heterogeneous collection of miscellaneous surfaces, objects, colours and brightness. Berlyne (1966) drew attention to the fact that humans and other animals spend more time looking at arrays that were irregularly arranged, had more components, had heterogeneous components, were irregularly shaped, were incongruous, were asymmetric, or were random – in other words, complex arrays, which, he argued, were intrinsically more attractive than simple arrays. Within a visual array there may be forms that we want to pay attention to, and other forms that for the moment we do not need to pay attention to. People spend much of

their time searching through natural arrays of forms, and practical visual search problems in the natural world are analogous to searching for targets in a laboratory visual array.

Parallel Processing

Sometimes the number of items in the array does not matter. Donderi and Zelnicker (1969) found that the average decision time for "same-different" responses to an array of between 2 and 11 black – on – white geometric shapes, where "same" meant all the shapes in the array (circles, squares, semicircles or triangles) were the same, and "different" meant that one was different from the rest, did not vary with the number of shapes in the array. The visual extent of the array was the same, regardless of the number of shapes. This finding was confirmed later (Donderi & Case, 1970) with arrays of between 2 and 14 shapes, and with either one shape or three shapes different from the rest. The search required to make a simple classification decision about an array of geometric forms does not take longer because there are more forms in the array.

Textons and Features

Array textures can be more dense than the few forms studied by Donderi and his students. Julesz and Bergen (1983) analyzed the structure of dense arrays of simple visual forms like Ls, Ts and Xs. Using a variety of experimental techniques they found that some array texture elements could easily be distinguished from one another, so that they formed easily segregated perceptual areas. For example, using exposure durations too short for eye movements, a randomly oriented L was easily detected in an area filled with randomly oriented Xs, but the same randomly oriented L was not easily detected among randomly ordered Ts. Julesz called the clear discrimination of certain textures preattentive discrimination. Julesz argued that the visual system segregates certain visual characteristics at a very early stage in visual processing, and that no conscious direction of attention across the visual field is required to discriminate among those perceptually segregated stimulus characteristics, like the difference between Ls and Xs, while consciously directed attention is required to find and report differences among stimuli bearing characteristics that the early visual system does not segregate. Julesz called his basic segregated textural units "textons."

Treisman and Gelade (1980) theorized that the basic perceptual components of a visual array corresponding to Julesz's textons were features like form (letter shape) and colour. If a visual search task required a decision based on the presence or absence of only one feature, for example, colour, then the decision would be fast and would not depend on the number of forms in the array. This was described as "pop-out," a characteristic of search using a single basic perceptual component. If a search task required a decision based on the presence or absence of a combination of basic features, for example a green H among green Os and red Hs, then the decision would be slow, it would take longer in the negative case (conjunction missing) and decision time would increase as the number of forms increased. This type of conjunction search was thought to be serial, because it required attention focused on each form in succession, and self-terminating, because it ended when the target was found, even if all of the forms had not been searched.

Julesz's and Treisman and Gelade's theories generated twenty years of research intended to specify what aspects of a visual array could be considered fundamental features, meaning features that would "pop out" if a search decision was based on their presence or absence. As a result, many such features have been found, including line orientation, luminance, colour, curvature, number, kind of line terminations, line intersections, figure closure, flicker, motion direction, binocular lustre, stereoscopic depth, apparent dimensionality and apparent lighting orientation (Northdurft, 1994; Enns, 1990; Wolfe, 1998) This means that if a distinction among forms in an array is made on any one of the bases described above, a target form can be found among the others in very little time, irrespective of how many such forms there are in the array.

Treisman (1982) also showed that physical grouping of the forms in the stimulus array significantly decreased decision times under the conjunction search conditions, which she explained as showing that the focus of conscious attention required by feature conjunction search could be modified by providing conditions under which many forms (all of those in a single grouping, for example) could be attended together.

What Kind of Search?

The serial – parallel difference based on conjunction versus simple feature search is weakened by the fact that decision time functions of the number of items in a search array do not determine whether search is serial or parallel, or whether search is self-terminating or exhaustive. Snodgrass and Townsend (1980) show that many possible combinations of serial and parallel search, with or without self-termination, can lead to identical decision time functions of the number of array items. Another weakness of a strict serial-parallel distinction was demonstrated in the analysis of over one million trials of search data from about 650 different observers (Wolfe, 1998), which produced a positively skewed unimodal distribution of search slopes (msec/item) and no clear distinction between slopes for serial and parallel search types. The entire data set also shows that the expectation from self-terminating serial searches; that is, the slope in msec/item will be twice as large when the target is absent as when it is present (because, on the average, twice as many items must be searched in an exhaustive serial search) is not met: the target absent/target present slope ratios were close to, but significantly larger than, 2.0. Donderi (1983, 1985) obtained similar results from "same-different" searches Over a series of experiments conducted with different numbers of letters randomly distributed over the display, target – absent decision times were about twice the target – present times at varying numbers of targets, but both same and different decision times were a negatively accelerated function of the number of letters in the display. Increasing target density (Donderi), as well as providing target grouping (Treisman) clearly has an accelerating effect on search rate. On the other hand, "natural" search – search carried out under conditions more closely approximating those of the real world, may make it harder to locate the information necessary to spatially locate the features that need to be analyzed to find a conjoint target (Wolfe, 1994), and as a result, search for targets in "natural" environments may be slower than laboratory experiments would predict.

Conspicuity and Discrimination

The efficiency with which one form in a visual scene can be attended is called its conspicuity (Engel, 1970, 1977). Conspicuity varies with both the characteristics of the form and the differences between the form and the background. A conspicuous form can be identified farther away from the point of visual fixation than an inconspicuous form. Engel pointed out that, because conspicuous forms can be identified over a wider visual field, fewer fixations over a given area are required to locate a conspicuous form than to locate an inconspicuous one.

The idea that search decision time is a function of the discriminability or conspicuity of a particular form, and that this provides as good an explanation as the difference between simple feature search and conjunction search, was pursued by Eckstein (1998) who tested the idea that conjunction targets are less accurately identified because the observer is dealing with a signal detection problem on two noisy dimensions, and response uncertainty on the two dimensions explains the added difficulty, either in terms of accuracy or speed, for conjunction search. Eckstein's experiments eliminated confounding factors like differences in peripheral sensitivity, because all of the forms were presented on a circle equidistant from fovea. The forms were ellipses with different contrasts and different orientations. The feature search was for an ellipse with either a different contrast or a different orientation, while the conjunction search was for, for example, a tilted high-contrast ellipse amidst tilted, straight, and both high and low-contrast ellipses.

Eckstein showed that computations based on signal detection theory differences between the feature and conjunction searches explained the obtained results better than a model based either on serial search alone, or a hybrid serial model with partial sampling. His conclusion was that simple discriminability differences explain the performance differences between feature and conjunction searches: conjunction searches are harder because the visual system has to resolve uncertainties on two characteristics of the visual input, rather than just one characteristic. Verghese and Nakayama (1993) also found that discriminability aspects of the sensory input, rather than capacity limitations on attention of the kind suggested by serial-search models, better predict the data from carefully controlled visual search experiments.

Studies of the individual properties of target forms show that differences between the target and distractor forms in luminance and in colour are both important determinants of visual conspicuity, and therefore of visual search time. If the chromaticity or luminance differences are great enough, the targets are sufficiently conspicuous so that the change in search time as a function of the number of distractor forms becomes very small (Nagy & Sanchez, 1992).

Eye Movements and Conspicuity

The study of conspicuity has been associated with the study of the number of visual fixations required to locate and identify a target. In an experiment involving searching for one of a previously cued set of equally conspicuous computer-generated forms, the eyes started from a central fixation point, moved to a point more or less at the centre of gravity of the area containing the forms to be searched, and then moved close enough to the target form on the third fixation for the target to be identified. The first fixations were not precisely on target, but they clearly responded to the layout of forms on the screen to be searched (Zelinsky et al., 1997).

Since the location of successive saccades was weighted by the location of the forms, Zelinsky et al. suggest that " The process by which gaze gradually converges on the target suggests a binary search strategy, rather than a sequential item-by-item search. The analysis occurring after the initial centre-of-gravity saccade effectively divides the display in two, isolating the hemifield in which the target is located. The second saccade then brings gaze to the centre of an object configuration on the selected side of the display, after which another binary decision is made". There was even "averaging" of position, and imperfect first saccade location, when there was only one item.

Bekkering and Neggers (2002) showed that the direction of a visual saccade in a search task depends on the task to be carried out. Observers had to either point to, or grasp between thumb and forefinger, a small block that was set on a flat horizontal surface. The vertical target was angled either 45 deg or 135 deg from the midline, and was either green or orange. There were 3, 6, or 9 distractor forms located near the target block that differed from the target block either in colour, in orientation, or both. Observers received cues for target colour and orientation before each trial, and all of the blocks – targets and distractor – were invisible in the darkened experimental room until the beginning of each trial. Once the observer had fixated a luminous reference point at the beginning of the trial, the target and distractor blocks were illuminated, and the observer either pointed at, or grasped, the target block depending on the task on that set of trials. Locations of target and distractor blocks were randomized from trial to trial.

The first saccade reached the target block just as often under the grasp as under the point condition, but the saccadic latencies were longer for grasping than for pointing, and latencies increased with the number of distracters. When the first saccade landed on a distractor (an error), the error could have been towards an object with a colour that matched the target (colour error) or towards an object whose orientation matched the target (orientation error). Over the entire experiment, there were more orientation than colour errors. The colour errors were equally distributed across the point and the grasp conditions, but the orientation errors were concentrated in the pointing condition. The authors regard this as evidence that visual search eye movements are influenced not only by the location of a target but also by what is to be done with it. The reduction of orientation errors in the grasp condition means that visual attention was influenced not just by target-distractor similarity, but by the response aspects of the task itself.

The Hick-Hyman Law

The first purpose of Shannon's communication theory was to quantify the amount of information transmitted from the message intended by the sender to the message that was actually received. The ideal communication channel transmits all information perfectly, so that every symbol received is the same symbol, in the same order, as the symbol sent. But communication channels are not perfect. During the Second World War when communications theory was being developed, the practical example of a communication channel was Morse code telegraphy. A Morse dot might get corrupted by static into a dash; a burst of random noise from lightning discharge might make it impossible to receive several successive symbols, and as a result the pattern of dots and dashes received might bear only a partial resemblance to the pattern sent. Shannon information theory provided a measure of the amount of information in the message sent, the amount transmitted between the message sent and the message received, and the amount in the message received. Because of transmission errors, the information in the message received might be different from the information in the message sent. The information in the message sent is called H(x), the information in the message received is called H(y), and the information common between the two messages – the transmitted information – is called T(x:y) (McGill and Quastler, 1954) If transmission is not perfect, then not all of H(x) gets transmitted, and T(x:y) is less than H(x). The difference between T(x:y) and H(x) is the information in the sent message that is not transmitted to the received message, and is therefore independent of it. It is called $H_y(x)$. If transmission is not perfect, then some of the information in the received message may also be independent of the information in the sent message, and this information will be called $H_x(y)$.

In the visual perception interpretation of information theory a situation can arise where an observer is not required to respond to all of the information that is made available over a communication channel. Consider the observer as a communication channel, the visual input as the sent message, and the observer's response to the visual input as the received message. Consider that the visual input in a specific experiment consists of equally likely presentations of a large red square, a small red square, a large blue square and a small blue square. The instructions require the observer to press one key when a red form is presented and another key when a blue form is presented. Using conventional information measurement, the ensemble of four equiprobable stimuli conveys $\log_2 4$, or 2 bits, of information. The ensemble of two equiprobable responses conveys $\log_2 2$, or 1 bit of information. In order to carry out the task perfectly, the communication channel (the link between observer input and output) only has to carry one bit of information: the binary decision as to whether a form is red or blue, rather than the quaternary decision, whether a form is large or small, red or blue. Size is, for the purposes of this experiment, and irrelevant dimension, and the observer need pay no attention to it in order to successfully complete the task. The transmitted information is only one bit. On the other hand, if the observer is required to make a different response to each form, then there will be four responses conveying $\log_2 4$, or 2 bits of response information, and the information about all four visual forms will be transmitted from the visual input to the response output.

Under these conditions, response performance efficiency measured by reaction time follows the Hick-Hyman law,

$RT = a + b*T(x:y)$

and so would be faster in the two-response situation than in the four-response situation. It should be noted in passing that T(x:y) depends on the relative frequencies (probabilities) of the visual stimuli and responses, and not just on how many of each there are.

RT can also be influenced (although this is not directly relevant to form perception) by the particular set of responses chosen, as well as the uncertainty implicit in the chosen set (Fitts & Switzer, 1962). There are also practiced tasks where no reduction in RT occurs as the number of response alternatives is reduced. But the conclusion to be drawn in demonstrating that the Hick-Hyman law applies to information processing of visual forms is that task requirements are an important determiner of how visual information is processed. An example from studies of perceptual arrays is the distinction (Treisman, 1982) between feature searches. If a conjunction search stimulus – say a green X among red Xs and green Os – is presented, and the task is simply to do a feature search – to indicate whether there are green items present, the decision time will be considerably less than if the observer must find the green X, because, in part, the response to the green X will occur less frequently than the response to either the red or green items. RT may still be longer than if the anomalous stimulus was not present, because the observer will respond both to the information in the input and the information transmission demands, but RT will be much shorter than if a true conjunction search is required.

Summary: Array and Form Complexity

Visual attention -- the ability to extract information from the visual display -- can be directed to different parts of the display even without changing visual fixation. The "conspicuity area" -- the area of the visual field in which a particular target can be identified -- varies with the conspicuity of the target. Real-world search for a target among distracters involves multiple fixations; not on successive targets, but on areas of the visual field that are near a group of targets, with successive refinements in fixation as the critical target is approached. The last fixation in a display with a target is considerably longer than the earlier fixations or than the fixation durations of displays without targets (Donderi, 1984).

Visual search of an array of forms cannot be strictly classified as either parallel – all forms searched at once, or strictly serial – each form searched in turn. Instead, the efficiency of a search as defined by the time consumed per form depends on a complex interaction between the number and spacing of forms present, the discriminability on one or more basic dimensions (colour, shape, size) of one form from another, and on the task itself. Differences in the distribution of forms over an array that are based on simple visual properties like colour or geometric shape are easily identified, but subtler differences that depend on a synthesis of different features will be harder to detect,

depending on what features are involved, what forms are involved, and how they are distributed over the visual field.

It is possible to measure the complexity of a single visual form based on the theory that the visual system analyzes visual input to generate a percept of minimum complexity. It is also possible to measure the probability of a single visual form based on the theory that the visual system generates a percept of maximum probability. Algorithmic information theory shows that the two criteria are the same.

The visual system also extracts information from form arrays by the rapid segregation of a long list of simple perceptual features, and by slower segregation when it is required to discriminate among more complex combinations of perceptual features. The time taken to extract the information depends in varying measures on the task, on the individual forms in the array, and on the form density of the array. Data about the perception of individual forms and data about the perception of form arrays must be linked to data and theories about the visual system in order to produce a useful and coherent understanding of the perception of forms and the perception of form arrays.

Perceptual Learning and Neural Circuits

Understanding form perception and the extraction of information from form arrays is aided by a theory of hierarchically organized perceptual learning. The visual system is hierarchically organized: cortical receptive fields exhibit successively more complex properties as they move further away from visual input: e.g. concentric circles, oriented contrast-specific line segments, stopped lines, moving lines over larger receptive fields, etc, (Kandel et al., 2000, pps. 492 – 506, 548 – 571). Ahissar (1997, 1999) proposed that perceptual learning begins with the activation of the most general level in the perceptual hierarchy, close to that with which we normally interact with the world. If a perceptual task cannot be resolved at the most general level – for example, by identifying it simply as a "triangle" or a "circle" or "red", then neurons are recruited at a more peripheral, and therefore more specific level, to accomplish the required task. With respect to the "pop-out" of features from an array, those features that are easily segregated – Julesz's "textons" or Treisman's simple shapes or simple colours – will be represented at the most general cortical level. But those feature combinations that take time to find – Treisman's feature conjunctions or the Ts and Ls of Julesz's arrays, will require analysis by more specific sets of more peripherally located neurons. This process is facilitated when practice on the easier (higher) level facilitates feedback to the peripheral neurons which represent details at the harder (lower) level, so that training on an easy version of a discrimination task facilitates learning that task far more effectively than continual practice on the hardest part of the task: a fact which, as Ahissar points out, was first established by Pavlov. This prediction from Ahissar's theory has been verified in practical studies of perceptual learning (Miller & Miller, 1968; McDonald, Martin, Williams & Hardy, 1973; McDonald & Martin, 1974). A corollary to this theory is that lower-level neurons will be recruited for a perceptual learning task only if they are needed because the neurons stimulated at a higher level by the sensory input cannot by themselves resolve the discrimination required by the response contingencies of the task.

Neural Circuit Theory

Ahissar's general theory of perceptual learning is consistent with the principles of several neural perceptual learning models (Grossberg, 1980, 1995, Rao & Ballard, 1999). Inspired by the cell-assembly theory of D. O. Hebb (Hebb, 1949; Freeman, 1991).

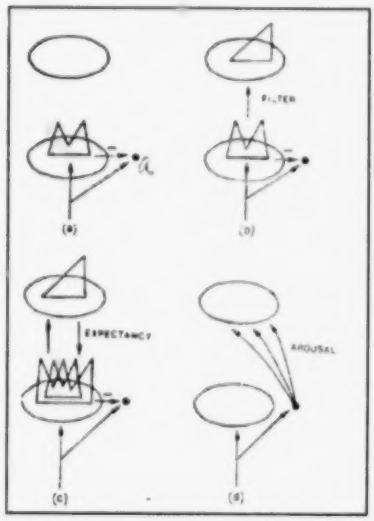


Figure 6. **Grossberg's neural circuit theory. Input (a) fires cells and inhibits arousal neurons (*a*). Neural code activates next level (b), which returns a mismatched signal (c), resulting in arousal which resets the synaptic weights, leading to new matching of input to second level. From Grossberg (1980).** (Copyright © 1980 by the American Psychological Association. Reprinted (or adapted) with permission.)

Grossberg's goal was to understand how a neural circuit shaped by synaptic plasticity that represented a percept could be stabilized in the presence of fluctuating sensory inputs and fluctuating attention. His solution was the concept of 'adaptive resonance.' The neural code representing a percept was expressed as synaptic weights linking cells at a peripheral level to cells at a higher level of the perceptual system. These cells were linked (as indeed almost all visual system cells are) to lower levels by feedback connections. If the firing pattern of the second-level neurons produced feedback to the first level that matched the first-level activity, then activity in the circuit would be potentiated and maintained, corresponding to a stable percept for as long as the activity persisted. If the feedback activity did not match the input-level activity, then an attentional mechanism

would provide a "reset" input to the cells in the circuit, allowing the synaptic weights to be recalculated. The new set of weights would produce different feedback, and when the feedback did match the input, the circuit would then be stabilized (Figure 6). Grossberg envisioned a hierarchy of analyzers of this type, and in fact has developed simulation models for many perceptual and learning phenomena based on the adaptive resonance principle (Cohen & Grossberg, 1987).

Rao and Ballard adopt Grossberg's hierarchy of feedback-based perceptual feature detectors, and suggest that the higher-level detectors are more general than lower-level detectors in both spatial coverage and in perceptual generality. Their approach is similar to the Grossberg theory in that their feedback consists of error-correcting signals that specify the mismatch between the lower-level and higher-level detector units (Figure 7). The receptive fields of primary visual cortex are organized to efficiently compress the incoming visual information (Olshausen & Field, 2000). Rao and Ballard assume that feedback feature detectors play a role in producing the efficient low-level feature detectors and also assume that these feature detectors, as well as those at higher levels, can be re-organized as the organism's needs or environment change, an assertion consistent with the Ahissar perceptual learning theory as well as other neural circuit theories (Law, Bear & Cooper, 1993).

The common property of these neural theories is that the cortical activation that corresponds to the experience of a percept occurs in a series of hierarchically organized feedback circuits, each circuit being activated by, and activating, circuits at both lower and higher levels. The connection weights of neurons linking circuits at one level to circuits at another are modified to produce a consistent firing pattern that links the higher-order circuits to lower-order neurons. The constraints on the organization of the circuits are ultimately the sensory input that fires the lowest-order neurons, and the response contingencies that reinforce the output of the entire hierarchy of active circuits (Milner, 1996).

Ahissar trained observers to improve their ability to "pop-out" visual features by systematic perceptual training that started with easy perceptual tasks and then advanced to harder tasks, a task order that is consistent with the learning order in his hierarchical model. The form complexity coding theories derived from Solomonoff's Algorithmic Information Theory are also consistent with Ahissar's hierarchical perceptual learning model and with the hierarchical neural models. The simpler, most probable and least differentiated form is one that is analyzed at a higher level in the perceptual hierarchy. It corresponds to a very general percept like "triangle" or "red" or perhaps even "car". Specific sets of clearly differentiated lower-level neurons are not directly implicated in the perception of "triangle", so any feedback to and from a large range of lower-level neurons will match the higher-level circuit sufficiently well so that the higher-level circuit alone is implicated in producing behaviour based on "triangle." In other words, the central nervous system does not specify higher-level forms in terms of lower-level neurons; it leaves the lower-level neurons either inactive or at unspecified levels of activity while it is activating the higher-level circuit corresponding to the general form. Such a form is more probable because its activation is consistent with the activation of a

larger number (a larger probability set) of lower-order neurons, and it is simpler because it requires the specification only of one higher-level circuit, rather than a narrowly specified hierarchy of several levels of neural circuits.

The more complex and less probable form is one that requires differentiation at multiple levels of the perceptual learning and visual system hierarchies. It is at the same time a "triangle" and "textured" and "small" and "pointed to the left." The specific multiple properties engage neurons at both the central and the peripheral levels of the visual system. If this more completely specified triangle is to be discriminated from other triangles, all levels of the neural feedback circuits must be equally well defined and activated as part of the form. As a result, the small, left-pointing, textured triangle demands a more precisely tuned circuit at every visual cortical level, and so engages more visual cortex neurons in the circuit activity required to respond correctly to this input. It is therefore a less probable form simply because it requires a greater specificity of activation in visual cortex neurons, (a smaller probability set) and is therefore less likely to be activated than is the general "triangle" form which has a larger set of potential activators.

An important prediction of Ahissar's theory is that it will take longer to learn tasks that require more peripheral visual circuits to be engaged, and that, when learned, these tasks will be harder to do, as measured by error rate, decision time, or an appropriate combination of the two. A second prediction is that perceptual learning will be faster if the higher-level circuits are engaged first, because these circuits feed down to the lower-level circuits and help to activate the appropriate circuits during later and more difficult stages of the perceptual learning process.

Colour and Linear Separability

One characteristic of form discrimination in a search array is called linear separability (Hodsoll & Humphreys, 2001; Bauer, Jolicoeur & Cowan, 1996). Bauer and McFadden (1997) showed that serial visual search for one of a set of three targets forms (circle, square or triangle) was speeded when each form was redundantly coded with a colour. Then they showed that a "linearly separable" target colour, defined as colours that occupied a corner, rather than the interior, of a triangle of six colours located in CIE 1976 colour space, produced faster search times than a target colour drawn from the interior of the colour triangle, where it could not be separated from the distractor colours by a line bisecting the colour space. The explanation based on Ahissar's perceptual learning theory is that the separation of colour features can be done at the highest level of generality by partitioning the opponent-colour space of red-green and blue-yellow vectors into two contiguous regions. Finer partitions, requiring what in Treisman's terms would be a conjunction of colour dimensions within that space, would require more neural circuits to become involved, and would therefore become more specific less probable and harder to learn.

A Synthesis of AIT, Perceptual Learning Theory and Neural Circuits

The code length, which is a property of AIT, when applied to visual forms, is commensurate with the depth of engagement of visual cortex neurons in analyzing a form. The simpler form that activates only a higher-level circuit has a shorter code, which corresponds to the simpler activation of fewer higher – order circuits. The more complex form that activates circuits at more levels of the visual system hierarchy has a longer code which corresponds to the greater specificity of the neural circuits which the more complex form activates.

Form probability is associated with form simplicity, and both are inversely related to AIT code length. According to Ahissar's perceptual learning theory, simpler forms activate higher-order neural circuits and they recruit fewer visual system neurons, leaving more of them in unspecified activity states. The circuit corresponding to a simpler form can be activated by a larger range of lower-level neurons, and so is more likely to be activated. The neural equivalent of its shorter AIT code length is the simpler higher-level circuit. Perception of a more complex form corresponds to the specification of circuits at more levels in the visual system hierarchy, which means that it can be activated by fewer input neurons and so is less likely to be activated. The neural equivalent of its longer AIT code is the greater specificity and larger number of its activated circuits.

Complexity of Real Displays

The information provided by a single display has to be evaluated with respect to all of the information needed to perform the task that the display is designed to serve, and complexity is not the only display characteristic related to how well a display fills that role. For one thing, the amount of information necessary to carry out the task puts a lower limit on the information complexity of the display. Then display designer must consider the costs and benefits of partitioning the required information across separate displays, which gains display simplicity but loses immediate accessibility, versus designing a single, always visible, but more complex display. Some recent evidence (Burns, 2000, Figure 8) suggests that putting more information on an already complex display can improve performance in a complex monitoring task.

The display medium is also an important consideration in predicting display performance. Ziefle (1998) carried out a series of well-controlled experiments in which she found better performance for many simple visual array search tasks (proofreading, searching lists for specified characters) on higher resolution computer screens, and even better performance on even higher-resolution paper. Low-resolution screens also produced symptoms of visual fatigue and effort. All of these tasks required discrimination of fine detail throughout the visual system hierarchy, and low-resolution input provided less resolution to visual feedback circuits in the periphery, and therefore made it difficult for the visual system to distinguish between similar peripheral feedback circuits encoding vital visual distinctions such as the difference between an h and a b, or an i and a j.

The organization and configuration of display information is as important as the total amount of information on the display. Much current display design research is about how to display a given amount of information, and the investigations focus on what kind of visual analogues should be used for numerical variables, how those variables should be combined, and how the display organization relates to the mental model of the task being controlled (Vicente et al., 1996; Carswell & Wickens, 1996; Bennett, Toms & Woods, 1993). A single-valued measure of display complexity provides no information about the way display information is configured, and so cannot answer the questions posed by research on separable versus integral dimensions, the proximity compatibility principle, configural displays and ecological displays.

Level of training will also interact with display complexity and task efficiency. The complex interacting levels of feedback circuits activated by a complex visual form or array require time to learn, but by organizing training based on Ahissar's perceptual learning principles, they can be learned efficiently. An experienced display user will be better at dealing with a given level of display complexity than an inexperienced user, and system designers have to consider the tradeoffs between the information required to do the task, the kind of training required to use a display, and the complexity of the display.

Summary

The measurement of visual form and visual array complexity evolved from an effort driven by gestalt psychology concepts of "good form," dating from the early years of the twentieth century, to a process that can now be understood through the convergent and complimentary analyses of algorithmic information theory, perceptual learning theory, and neural circuit theory. Simple forms and arrays activate generalized perceptual learning processes corresponding to higher-level neural circuits: they are coded more simply in algorithmic information theory, and their simpler code corresponds to simpler neural circuitry activated by a wider range of more probable inputs. Complex forms and arrays activate more specific perceptual learning processes, corresponding to central and peripheral neural circuits: their more complex codes in algorithmic information theory correspond to more complex neural circuitry that is activated by a smaller range of less probable inputs.

Complexity is a quantifiable property of a visual display. Its lower limit is fixed by the information required to complete the tasks served by the display. Once the display information requirements have been fixed and the physical design limits have been specified, it the display designer's job to arrange and sequence the information to suit the task demands and to meet other criteria of efficient display design: for example, proximity compatibility (Wickens and Hollands, 2000, pps 97 – 100), integrality versus separability (Bennett, Toms & Woods, 1993), or information dimension selection (Carswell, 1992, Carswell & Wickens, 1996). But independent of other criteria, high display complexity, measured by compressed file size, is correlated with higher subjectively rated display complexity, increased target search time and decreased target search accuracy for several types of complex visual displays (Donderi, 2003).

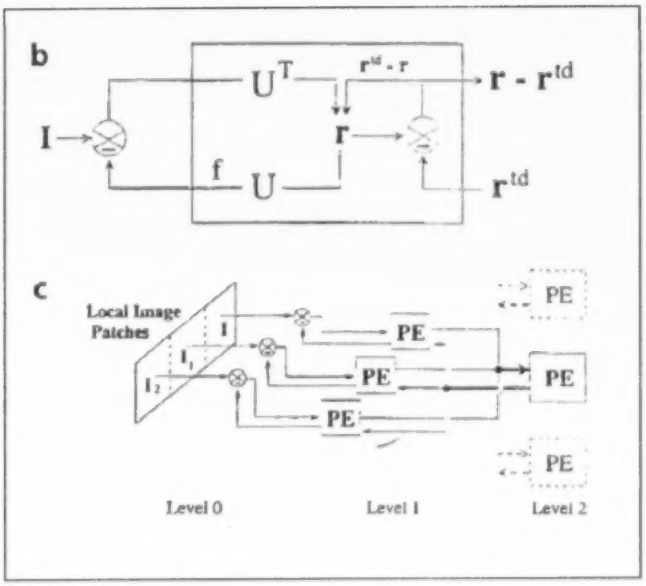


Figure 7. **A feature detector (b) and a hierarchy of feature detectors (c) from Rao and Ballard (1999).** (Reprinted with permission. Copyright Nature Publishing Group, 1999)

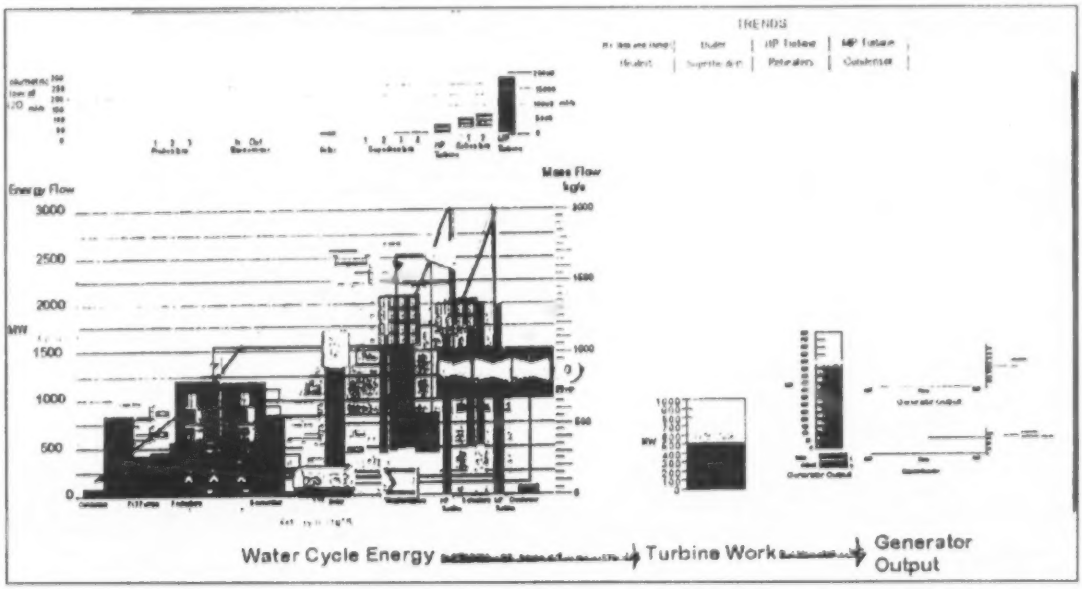


Figure 8. **Example of a complex display from Burns (2000) which produced the best fault diagnosis performance**. (Reprinted with permission from *Human Factors*, Vol. 42, No. 2, 2000. Copyright 2000 by the Human Factors and Ergonomics Society. All rights reserved.)

References

Ahissar, M. (1999). Perceptual Learning. *Current Directions in Psychological Science*, 8(4), 124 – 128.

Ahissar, M. & Hochstein, S. (1997). Task difficulty and the specificity of perceptual learning. *Nature*, 387, 401 – 406.

Anderson, N.S. & Leonard, J.A. (1958). The recognition, naming and reconstruction of visual figures as a function of contour redundancy. *Journal of Experimental Psychology*, 56(3), 262 – 270.

Arnoult, M.D. (1960). Prediction of perceptual responses from structural characteristics of the stimulus. *Perceptual and Motor Skills*, 11, 261-268.

Attneave, F. (1959). *Applications of information theory to psychology: A summary of basic concepts, methods and results*. New York: Holt, Rinehart & Winston.

Bauer, B. & McFadden, S. (1997). Linear separability and redundant colour coding in visual displays. *Displays*, 18, 21 – 28.

Bauer, B., Jolicoeur, P. & Cowan, W.B. (1996). Visual search for colour targets that are or are not linearly separable from distractors. Vision Research, 36, 1439 – 1465.

Bekkering, H. & Neggers, S.F.W. (2002). Visual search is modulated by action intentions. *Psychological science*, 13(4), 370 – 374.

Bennett, K.B., Toms, M.L. & Woods, D.D. (1993). Emergent features and graphical elements: Designing more effective configural displays. *Human Factors*, 35(1), 71 – 97.

Berlyne, D.E. (1966). Curiosity and exploration. *Science*, 153, 25 – 33.

Boring, E.G. (1959). *A history of experimental psychology*. New York: Appleton-Century.

Boring, E.G. (1942) *Sensation and perception in the history of experimental psychology*. New York: Appleton-Century.

Bourne, L.E. & Haygood, R.C. (1959). The role of stimulus redundancy in concept identification. *Journal of Experimental Psychology*, 58, 232 – 238.

Burns, C.M. (2000) Putting it all together: Improving display integration in ecological displays. *Human Factors*, 42(2), 226 – 241.

Carmichael, L., Hogan, H.P. & Walter, A.A. (1932). An experimental study of the effect of language on the reproduction of a visually perceived form. *Journal of Experimental Psychology*, 15, 73 – 86.

Carswell, C.M. (1992) Choosing specifiers: An evaluation of the basic tasks model of graphical perception. *Human Factors*, 34(5), 535 – 554.

Carswell, C.M. & Wickens, C.D. (1996). Mixing and matching lower-level codes for object displays: evidence for two sources of proximity compatibility. *Human Factors*, 38 (1), 1 – 22.

Chater, N. (1996). Reconciling simplicity and likelihood principles in perceptual organization. *Psychological Review*, 103 (3), 566 – 581.

Cohen, M.A. & Grossberg, G. (1987) Masking fields: a massively parallel neural architecture for learning, recognizing and predicting multiple groupings of patterned data. *Applied Optics*, 26, 1866 – 1891.

Donderi, D.C. (2003) *A complexity measure for electronic displays: final report on the experiments.* Government of Canada: Department of National Defence: Defence Research and Development – Toronto. (Draft Report)

Donderi, D.C. (1984). Eye movements during visual search of random letter displays. Presented at the American Psychological Association convention, Toronto

Donderi, D.C. (1973). Changes in visual recall memory following discrimination learning. *Canadian Journal of Psychology*, 27(2), 210 – 219.

Donderi, D.C. & Case, B. (1970). Parallel visual processing: Constant same-different latency with two to fourteen shapes. *Perception and Psychophysics*, 9(5b), 373 – 375.

Donderi, D.C. & Zelnicker, D. (1969) Parallel processing in visual same-different decisions. *Perception and Psychophysics*, 5(4), 197 – 200.

Donderi, D.C. (1983). Acquisition and decision in visual same-different search of letter displays. *Perception and Psychophysics*, 33(3), 271 – 282.

Donderi, D.C. (1985). Acquisition and decision during target search of visual letter displays. *Perception and Psychophysics*, 37(4), 345 – 354.

Donderi, D.C. (1990). Response – contingent variation in visual recall: Evidence of a dynamic memory trace. *Canadian Journal of Psychology*, 44(4), 423 – 434.

Donderi, D.C., Case, B., Rossoff, L. & Willis, J. (1970). Change in visual memory following discrimination learning. *Canadian Journal of Psychology*, 24(6), 401 – 416.

Eckstein, M.P. (1998). The lower visual search efficiency for conjunctions is due to noise and not serial attentional processing. *Psychological Science*, 9(2), 111-118.

Engel, F.L. (1977). Visual conspicuity, visual search and fixation tendencies of the eye. *Vision Research*, 17, 95 – 108.

Engel, F.L. (1970). Visual conspicuity, directed attention and retinal locus. *Vision Research*, 11, 563 – 576.

Enns, J.T. (1990) Three-dimensional features that pop out in visual search. In Brogan, D. (ed.) *Visual Search*. London: Taylor and Francis.

Fechner, G.T. (1860/1966). *Elements of psychophysics*, Volume 1, tr. H. E. Adler, ed. D. H. Howes, E. G. Boring. New York: Holt, Rinehart and Winston.

Fitts, P.M. & Switzer, G. (1962). Cognitive aspects of information processing: 1. The familiarity of S-R sets and subsets. *Journal of Experimental Psychology*, 63(4), 321 – 329.

Fitts, P.M., Weinstein, M., Rappaport, M., Anderson, N. & Leonard, J.A. (1956). Stimulus correlates of visual pattern perception: a probability approach. *Journal of Experimental Psychology*, 51, 1 – 11.

Freeman, W.J. (1991) The physiology of perception. *Scientific American*, 276, 78 – 83.

Garner, W.R. (1962) *Uncertainty and structure as psychological concepts*. New York: Wiley.

Garner, W.R. (1970). Good patterns have few alternatives. *American Scientist*, 58, 34 – 42.

Garner, W.R. (1974). *The processing of information and structure*. New York: Wiley.

Gibson, J.J. (1929). The reproduction of visually perceived forms. *Journal of Experimental Psychology*, 12, 1-39.

Gibson, J.J. (1950). *The perception of the visual world*. Boston: Houghton Mifflin.

Gibson, J.J. (1955a) Parallax and perspective during aircraft landings. *American Journal of Psychology*, 58, 372 – 375.

Gibson, J.J. (1955b). The optical expansion – pattern in aerial locomotion. *American Journal of Psychology*, 58, 480 – 484.

Gibson, J.J. (1966). *The senses considered as perceptual systems*. Boston: Houghton Mifflin.

Goldstein, E.B. (1980) *Sensation and Perception*. Belmont, CA: Wadsworth.

Grossberg, S. (1980) How does a brain build a cognitive code? *Psychological Review*, 87(1), 1 – 51.

Grossberg, S. (1995). The attentive brain. *American Scientist*, 83 (5), 438 – 449.

Hanawalt, N.G. (1952). The method of comparison applied to the problem of memory change. *Journal of Experimental Psychology*, 43, 37 – 42.

Hebb. D.O. (1949). *Organization of behavior: A neuropsychological theory*. New York: Wiley.

Hebb, D.O. & Foord, E.N. (1945). Errors of visual recognition and the nature of the trace. *Journal of Experimental Psychology*, 35, 335 – 348.

Hochberg, J. & Brooks, V. (1960). The psychophysics of form: reversible-perspective drawings of spatial objects. *American Journal of Psychology*, 73, 624 – 628.

Hochberg, J. (1957). Effects of the gestalt Revolution: The Cornell symposium on perception. *Psychological Review*, 64 (2), 73 – 84.

Hodsoll, J. & Humphreys, G.W. (2001). Driving attention with the top down: the relative contribution of target templates to the linear separability effect in the size dimension. *Perception and Psychophysics*, 63 (5), 918 – 926.

Hubel, D.H. & Wiesel, T.N. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology*, 195: 215 – 244.

Julesz, B. & Bergen, J.R. (1983). Parallel versus serial processing in rapid pattern discrimination. *Nature*, 303, 696 – 698.

Kandel, E.R., Schwartz, J.H. & Jessell, T.M. (2000). *Principles of Neural Science*. New York: McGraw-Hill.

Kling, J.W. & Riggs, L.A. (1971). *Woodworth and Schlosberg's Experimental Psychology, 3rd ed.* New York: Holt, Rinehart & Winston.

Koffka, K. (1935). *Principles of gestalt Psychology*. New York: Harcourt, Brace & World.

Kohler, W. (1947), *Gestalt Psychology*. New York: Liveright.

Law, C.C., Bear, M.F. & Cooper, L.N. (1993). Role of the visual environment in the formation of receptive fields according to the BCM theory. *Brain Research*, 102, 287 – 301.

Leeuwenberg, E.L.J. (1968). *Structural information of visual patterns*. The Hague: Mouton de Gruyter.

Leeuwenberg, E.L.J. (1969). Quantitative specification of information in sequential patterns. *Psychological Review*, 76(2), 216 – 220.

Lockead, G.R. (1966). Effects of dimensional redundancy on visual discrimination. *Journal of Experimental Psychology*, 72(1), 95 – 104.

MacKay, D.M. (1962). Theoretical models of space perception. In Muses, C. A. *Aspects of the theory of artificial intelligence*. New York: Plenum, pps. 83 – 103.

Mavrides, C.M. & Brown, D.R. (1969). Discrimination and reproduction of patterns: Feature measures and constraint redundancy as predictors. *Perception and Psychophysics*, 6(5), 276 – 280.

McDonald, S. & Martin, G. (1974). A comparison of three errorless and two traditional methods of teaching discriminations, and their effects on generalization and retention with severely and moderately retarded girls. Paper presented at the American Psychological Association, New Orleans, LA.

McDonald, S., Martin, G.L., Williams, L. & Hardy, L. (1973). Errorless vs. traditional training on a simultaneous name discrimination task with severely retarded girls. Paper presented at the American Psychological Association, Montreal, Quebec.

McGill, W. & Quastler, H. (1954) Standardized nomenclature: an attempt. In Quastler, H (ed). *Information theory in psychology: Problems and Methods*. Glencoe, IL: Free Press.

Miller, A. & Miller, E.E. (1968). Symbol accentuation: the perceptual transfer of meaning from spoken to printed words. American Journal of Mental Deficiency, 73 (2), 200 – 208.

Miller, G.A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, 81 – 97.

Milner, P.M. (1996). Neural representations: some old problems revisited. *Journal of Cognitive Neuroscience*, 8(1), 69 – 77.

Murphy, G. (1949). *Historical introduction to modern psychology*. New York: Harcourt, Brace

Murray, E. (1908). A qualitative analysis of tickling: its relation to cutaneous and organic sensation. *American Journal of Psychology*, 19, 289 – 344.

Nagy, A.L. & Sanchez, R.R. (1992). Chromaticity and luminance as coding dimensions in visual search. *Human Factors*, 34(5), 601-614.

Northdurft, H-C. (1994). Common properties of visual segmentation. In Morgan, M. J. (ed). *Higher-Order Processing in the Visual System*, pps. 245 - 267. Chichester: Wiley.

Olshausen, B.A. & Field, D.J. (2000). Vision and the coding of natural images. *American Scientist*, 88, 238 – 245.

Quastler, H. (ed). (1954). *Information theory in psychology: Problems and methods*. Glencoe, IL: Free Press.

Rao, R.P.N. & Ballard, D.H. (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature neuroscience*, 2 (1), 79 – 87.

Rock, I. & Engelstein, P. (1959). A study of memory for visual form. *American Journal of Psychology*, 72, 221 – 229.

Rock, I. & Palmer, S. (1990). The legacy of gestalt psychology. *Scientific American*, 238, 84 – 90.

Shannon, C.E. & Weaver, W. (1949). *The mathematical theory of communication*. Urbana: University of Illinois Press.

Simon, H.A. & Kotovsky, K. (1963). Human acquisition of concepts for sequential patterns. *Psychological Review*, 70, 534 – 546.

Simon, H.A. (1972). Complexity and the representation of patterned sequences of symbols. *Psychological Review*, 1972, 79(5), 369 – 382.

Snodgrass, J.G. & Townsend, J.T. (1980). Comparing parallel and serial models: theory and implementation. *Journal of Experimental Psychology: Human Perception and Performance*, 6(2), 330 – 354.

Solomonoff, R.J. The application of algorithmic probability to problems in artificial intelligence. In Kanal, L.N. & Lemmer, J.F. (eds.). *Uncertainty in artificial intelligence*. North-Holland, Elsevier.

Sperry, R.W. (1958). Physiological plasticity and brain circuit theory. In Harlow, H. F. & Woolsey, C. N. (eds.) *Biological and biochemical bases of behavior*, pps. 401 – 424. Madison: University of Wisconsin Press.

Stevens, S.S. (1975). *Psychophysics: Introduction to its perceptual, neural and social prospects*. ed. G. Stevens. New York: Wiley, 1975.

Thomas, H. (1968). Spatial models and multidimensional scaling of random shapes. *American Journal of Psychology*, 81, 551 – 558.

Titchener, E.B. (1914). On the production of an artificial hiss. *Proceedings of the American Philosophical Society*, 53, 323-329.

Treisman, A. & Gelade, G. (1980). A feature – integration theory of attention. *Cognitive Psychology*, 12, 97 – 136.

Treisman, A. (1982). Perceptual grouping and attention in visual search for features and for objects. *Journal of Experimental Psychology: Human Perception and Performance*, 8(2), 194 – 214.

van der Helm, P.A. (2000) Simplicity versus likelihood in visual perception: from surprisals to precisals. *Psychological Bulletin*, 126(5), 770 – 800.

Verghese, P. & Nakayama, K. (1993). Stimulus discriminability in visual search. *Vision Research*, 2453 – 2467.

Vicente, K.J., Moray, N., Lee, J.D., Rasmussen, J., Jones, B.G., Brock, R. & Djemil, T. (1996) Evaluation of a Rankine cycle display for nuclear power plant monitoring and diagnosis. *Human Factors*, 38 (3), 506 – 521.

Wickens, C.D. & Hollands, J.G. (2000). *Engineering psychology and human performance*, 3rd ed. Upper Saddle River NJ: Prentice-Hall.

Wolfe, J.M. (1998) Visual search. In Pashler, H. (ed) *Attention*. London, UK: University College London Press.

Wolfe, J.M. (1994). Visual search in continuous, naturalistic stimuli. *Vision Research*, 34 (9), 1187 – 1195.

Wolfe, Jeremy M. (1998). What can 1 million trials tell us about visual search? *Psychological Science*, 9(1), 33-39.

Wolfram, S. (2002). *A new kind of science*. Champaign, IL: Wolfram Research

Zelinsky, G.J., Rao, R.P.N., Hayhoe, M.M. & Ballard, D. (1997). Eye movements reveal the spatiotemporal dynamics of visual search. *Psychological science*, 8(6), 448-453.

Ziefle, M. (1998) Effects of display resolution on visual performance. *Human Factors*, 40(4), 554 – 558.

DOCUMENT CONTROL DATA SHEET

1a. PERFORMING AGENCY
Human Factors North Inc., 118 Baldwin Street, Toronto, Ontario, M5T 1L6

2. SECURITY CLASSIFICATION
UNCLASSIFIED
Unlimited distribution -

1b. PUBLISHING AGENCY
DRDC Toronto

3. TITLE
(U) Visual complexity: A review

4. AUTHORS
Don C. Donderi

5. DATE OF PUBLICATION
March 4 , 2003

6. NO. OF PAGES
48

7. DESCRIPTIVE NOTES

8. SPONSORING/MONITORING/CONTRACTING/TASKING AGENCY
Sponsoring Agency:
Monitoring Agency:
Contracting Agency : DRDC Toronto
Tasking Agency:

9. ORIGINATORS DOCUMENT NO.
Contract Report CR 2003-047

10. CONTRACT GRANT AND/OR PROJECT NO.
W7711-017758/001/TOR

11. OTHER DOCUMENT NOS.

12. DOCUMENT RELEASABILITY
Unlimited distribution

13. DOCUMENT ANNOUNCEMENT
Unlimited announcement

14. ABSTRACT

(U) The definition and measurement of the complexity of visual forms and visual arrays is reviewed, starting with the contributions of gestalt psychology at the beginning of the twentieth century and ending with current research in algorithmic information theory, perceptual learning theory, and neural circuit theory. Complexity and probability are shown to be linked through their reciprocal relationship as developed in algorithmic information theory, which is shown in turn to be consistent with important aspects of perceptual learning theory and neural circuit theory. The emphasis is on applications to visual displays.

(U) On passe en revue la définition et la mesure de la complexité des formes visuelles et des groupements visuels, depuis les contributions du gestaltisme au début du 20e siècle jusqu'aux recherches actuelles sur la théorie de l'information algorithmique, la théorie de l'apprentissage perceptuel et la théorie du circuit neuronal. Il est démontré que la complexité et la probabilité sont interreliées dans le cadre de la théorie de l'information algorithmique, qui s'avère quant à elle compatible avec d'importants aspects de la théorie de l'apprentissage perceptuel et de la théorie du circuit neuronal. L'accent est mis sur les applications dans le contexte de la visualisation.

15. KEYWORDS, DESCRIPTORS or IDENTIFIERS

(U) complexity; visual displays; visual perception; human performance; information processing; algorithmic information theory; perceptual learning theory; neural circuit theory